Network-Scale Engineering: Systems Approaches to Synthetic Biology

Dissertation Advisor: Dr. Pamela A. Silver Patrick M. Boyle

Network-scale engineering: systems approaches to synthetic biology

ABSTRACT

The field of Synthetic Biology seeks to develop engineering principles for biological systems. Modular biological parts are repurposed and recombined to develop new synthetic biological devices with novel functions. The proper functioning of these devices is dependent on the cellular context provided by the host organism, and the interaction of these devices with host systems. The field of Systems Biology seeks to measure and model the properties of biological phenomena at the network scale. We present the application of systems biology approaches to synthetic biology, with particular emphasis on understanding and remodeling metabolic networks. Chapter 2 demonstrates the use of a Flux Balance Analysis model of the *Saccharomyces cerevisiae* metabolic network to identify and construct strains of *S. cerevisiae* that produced increased amounts of formic acid. Chapter 3 describes the development of synthetic metabolic pathways in *Escherichia coli* for the production of hydrogen, and a directed evolution strategy for hydrogenase enzyme improvement. Chapter 4 introduces the use of metabolomic profiling to investigate the role of circadian regulation in the metabolic network of the photoautotrophic cyanobacterium *Synechococcus elongatus* PCC 7942. Together, this work demonstrates the utility of network-scale approaches to understanding biological systems, and presents novel strategies for engineering metabolism.

Acknowledgements

During my time in the Silver lab, I have been fortunate to work with many smart, supportive, and entertaining friends and colleagues. I thank Pam Silver, my advisor, for fostering this environment, and allowing me to work independently and explore many different projects. I would also like to thank Jeff Way, who has served as a second mentor during my graduate work. There is no way to appropriately list all the graduate students and postdocs that I would like to thank; the entire Silver lab crew has consistently made the lab an entertaining and exciting place to do science. Caleb Kennedy introduced me to the project that eventually became Chapter 2 of this dissertation, and convinced me to combine computational and experimental approaches in my work. Christina Agapakis and Jake Wintermute likewise convinced me to combine explosive hydrogen experiments, blogging, and Youtube into my graduate experience. I also thank Devin Burrill, Mara Inniss, Bill Senapedis, Michael Moore, QingQing Wang, Danny Ducat, Dave Savage, Karmella Haynes, Noah Taylor, Camille Delebecque, Bruno Afonso, Zeev Waks, Buz Barstow, Gerald Grandl, Matt Mattozzi, and Mathias Voges for being friends and scientific collaborators. Two talented graduate students, Kelly Brock and Renan Escalante-Chong, worked with me when they were undergraduates and helped me with computational and experimental work.

I was also able to work with many people outside of the Silver lab on a number of projects. I owe a lot to the Harvard University Center for the Environment, especially Dan Schrag, Brenda Hugot, and Kellie Corcoran, for financial support and for providing excellent classes and seminars on the issues and challenges surrounding climate change. Likewise, the NSF Synthetic Biology Engineering Research Center, especially Megan Palmer, Kevin Costa, Leonard Katz, Genya Dana, and Drew Endy, provided financial support and supported work on the environmental and social consequences of synthetic biology. The 2010 Harvard iGEM team, including Aaron Deardon, Jonathan DeWerd, Alex Gedeon, Jackie Quinn, Morgan Paull, Anu Raman, Mark Thielmann, Lu Wang, Julia Winn, Oliver Medvedik, Alain Viel, Tamara Brenner, Kurt Schellenberg, Sarah Mathews, and Jagesh Shah, made a great team and got me into plant biology. John Asara and the entire Beth Israel Mass Spectrometry Core were great collaborators and made the metabolomics work in Chapter 4 possible. Kathy Buhl and the Systems Biology Lab Ops team have dealt with frequent and strange reagent requests, and kept the lab running well in the process. In the BBS Office, Kate Hodgins, Maria Bollinger, Steve Obuchowski, and Daniel Gonzalez were a great resource and kept me on track to graduation. In the Systems Biology Department and at the Wyss Institute, Leah VanVaerenewyck, Van Ferrell, and Becky Ward have been indispensible in assisting me with grant and fellowship applications.

I thank Fred Winston for his thoughtful advice, and for serving as the chair of my Dissertation Advisory Committee and my defense committee. Lew Cantley, Chris Marx, and Galit Lahav have also been very helpful members of my DAC. I also thank George Church, Ann Hochschild, and Kristala Prather for serving on my defense committee, and for providing feedback on this dissertation.

Finally, I would like to thank my friends for their support, and for driving me to be a better scientist. I thank the whole Boyle family in Massachusetts for their love, support, and delicious Thanksgiving meals. I thank my parents, Joseph and Aurora, and my brother Charles, for their love, for being adventurous, and for always being there for me. I thank my fiancée, Dr. Marissa Hamrick, for everything.

Table of Contents

List of Figures

List of Tables

Copyright Information

Chapter 1

Introduction

"The beauty of a living thing is not the atoms that go into it, but the way those atoms are put together. Information distilled over 4 billion years of biological evolution. Incidentally, all the organisms on the Earth are made essentially of that stuff. An eyedropper full of that liquid could be used to make a caterpillar or a petunia if only we knew how to put the components together."

— Carl Sagan

ON SYNTHETIC BIOLOGY[1]

As Synthetic Biologists, we study evolved biological systems to identify modular, reusable "parts," and repurpose these parts to design biological devices with new functions. Our primary goal, which has yet to be fully realized, is to establish engineering design principles for biological systems. Through quantitative modeling and measurement, as well as trial and error, a vast array of synthetic devices have demonstrated that the rational design of biological systems is possible (for additional reviews, see Drubin et al. 2007; Agapakis & Silver 2009; Boyle & Silver 2009; Haynes & Silver 2009; Boyle & Silver 2011; Burrill et al. 2011).

The ability to construct novel biological systems has been facilitated by the cost of two essential technologies: DNA sequencing and DNA synthesis (Figure 1.1). High throughput sequencing technology has resulted in the sequencing of thousands of genomes throughout the known biosphere. This has generated a vast library of biological parts that can be integrated into new devices (Bayer et al. 2009). In addition, sequencing of entire bacterial genomes is quickly becoming a routine and affordable process (Qin et al. 2010). DNA synthesis remains approximately 6 orders of magnitude more expensive than sequencing, but the synthesis of multi-kilobase DNA constructs is well within the resources of modern laboratories (Carlson 2010). DNA synthesis allows the design of constructs that are codon- and restriction site-optimized, a feature that would be difficult or prohibitively expensive to produce via polymerase chain reaction (PCR) and site-directed mutagenesis.

---

[1] Portions of this chapter were originally published in the following papers:

1. Boyle, P. M. & Silver, P. A. Harnessing nature's toolbox: regulatory elements for synthetic biology. *Journal of the Royal Society, Interface / the Royal Society* **6 Suppl 4**, S535–46 (2009)
2. Boyle, P. M. & Silver, P. A. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metab Eng* (2011).doi:10.1016/j.ymben.2011.10.003

The full text of these papers is included as Appendices B and C, respectively. Excerpts will be identified via footnotes in this chapter.

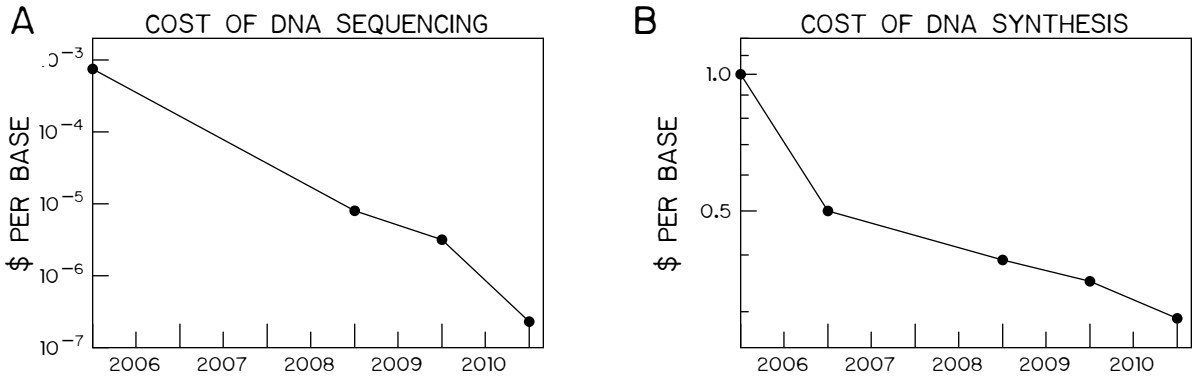


Figure 1.1 Cost of DNA sequencing and DNA synthesis during my time in graduate school
Source: (Carlson 2010 and personal communication with Rob Carlson)

While advances in DNA synthesis have made it possible to construct any desired DNA sequence, the utility of this power has been limited by the lack of predictive tools that can accurately model how a novel genetic construct will behave *in vivo*. Given the complexity of biological systems, it is unlikely that every aspect of an organism can be measured and understood prior to engineering it. However, biological design principles can be developed to abstract biological systems into comprehensible components.

BIOLOGICAL ABSTRACTION

Synthetic biologists typically abstract biological organization into three basic levels of complexity: parts, devices, and chassis (Endy 2005; Andrianantoandro et al. 2006). The divisions between these levels are perhaps less well-defined than in other engineering disciplines (Agapakis & Silver 2009), but parts, devices, and chassis roughly correspond to genetic elements, collections of interacting genetic elements, and host cells, respectively. For example, promoters, open reading frames, and transcriptional terminators are considered to be parts; a set of open reading frames that interact in a regulated manner (e.g. transcriptional regulation or a metabolic pathway) are considered

to be devices; and organisms with tractable genetic tools such as *Escherichia coli* and *Saccharomyces cerevisiae* are considered to be chasses.

Chassis engineering is central to most metabolic engineering efforts: a chassis is selected for a desirable metabolic property (such as producing large amounts of ethanol from glucose in the case of *S. cerevisiae*), and carbon flux to competing reactions is minimized to improve product yields. The engineered relationship between a device and its host chassis will vary depending on the application. Some applications require that a synthetic device functions "orthogonally" to the host chassis, acting independent of host cell regulatory systems. Along these lines, minimal cell chassis and orthogonal *in vivo* translation systems have been proposed, to minimize the interference of host cell functions with synthetic devices (Forster & Church 2006). Alternatively, devices may be engineered to integrate with their host chassis. For example, synthetic biologists can reconfigure a "maximal chassis," an organism with specific capabilities that can be leveraged by a synthetic device. Maximal chassis selection requires identifying genetically tractable species that most closely fit the desired application. Minimal chassis are less appropriate for industrial applications; they are not sufficiently flexible or robust for the conditions found in large-scale bioreactors (Keasling 2010). In fact, a variety of maximal chassis, such as *S. cerevisiaie* and *Pichia pastoris*, have been used to develop metabolic engineering strategies. We will briefly discuss emerging metabolic engineering strategies using synthetic biology approaches in a variety of chassis.

ENGINEERING METABOLISM WITH SYNTHETIC BIOLOGY[2]

Metabolic Engineering[3] can be considered as a specific application of Synthetic Biology: engineering the natural metabolism of a biological organism to produce a desired product (Boyle &

[2] This section is adapted from Boyle and Silver, 2011

Silver 2011; Keasling 2008; Carothers et al. 2009; Yadav & Stephanopoulos 2010). Natural metabolic pathways are controlled by myriad regulatory systems, for example transcription factors and promoters, that can be repurposed by synthetic biologists to modulate pathway components (Figure 1.2A). Ideally, a quantitative understanding of the transcription, translation, interactions, and kinetics of a metabolic pathway, as well as how that pathway interfaces with the host cell's metabolism, enables tuning of pathway components to maximize product yields (Figure 1.2B). In practice, our ability to tune pathways has improved as the fundamental principles of metabolism and biological regulation continue to be discovered.

Many synthetic regulatory devices to date have utilized elements of transcription, RNA processing, and translation to modulate device behavior (Boyle & Silver 2009). In the context of metabolic engineering, modifications to biological regulation are intended to maximize metabolic flux to the desired product. In most cases, this is accomplished via adjustments in enzyme expression levels, along with the elimination of competing pathways via gene knockout (Stephanopoulos 1999).

The structure and function of evolved metabolic networks suggests that this process of pathway optimization requires an understanding of how control is distributed across the entire pathway (Dekel & Alon 2005; Fell 1997; Zaslaver et al. 2004). In essence, pathway optimization is a multivariate problem, with no single "rate limiting step" to target. Furthermore, simple over-expression of pathway enzymes is often detrimental to product yields, through both the depletion of essential cellular reserves and the buildup of toxic metabolic intermediates (Alper, Miyaoku & Stephanopoulos 2005c; Jones et al. 2000; Raab et al. 2005). Efforts to model synthetic biological

---

[3] The term "Metabolic Engineering" predates the term "Synthetic Biology" (Keasling 2010). In this thesis I use Metabolic Engineering to refer to the process of redirecting metabolic flux to a desired product. Some reviews (such as Yadav & Stephanopoulos 2010) discuss Synthetic Biology and Metabolic Engineering as representing different design philosophies; I prefer to consider all relevant approaches to redirecting metabolic flux as Metabolic Engineering.

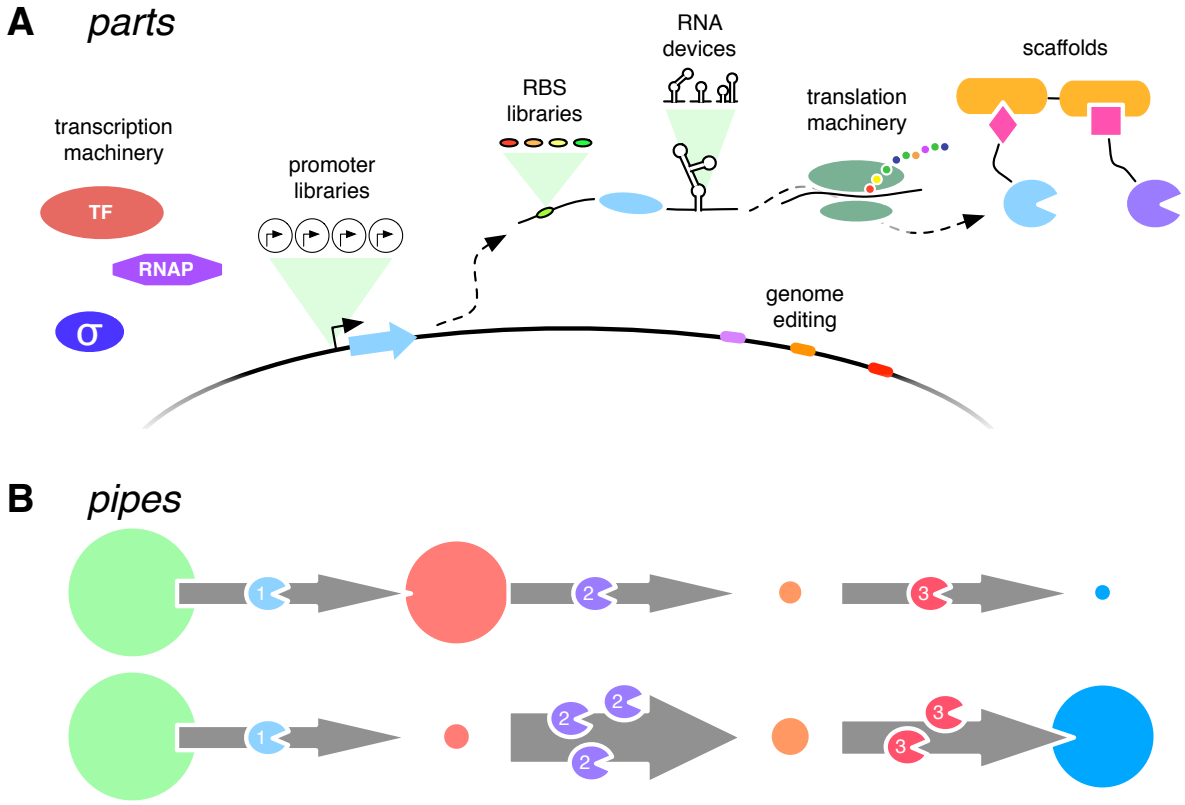


Figure 1.2 Combining synthetic biological "parts" to engineer metabolic "pipes." (A) Synthetic biologists use a variety of parts to adjust the functioning of metabolic pathways. Transcription machinery, enzyme promoters, ribosome binding sites (RBS), and translational machinery can be modified to adjust the concentration of an enzyme. RNA devices can modulate mRNA degradation and translation efficiency. Pathway enzymes can be assembled on scaffolds to optimize the spatial organization of a pathway. Genome editing approaches can be used to adjust host metabolism to improve flux through the target pathway. (B) A "pipe" of key pathway enzymes can be tuned to increase product titers. In this conceptual example, enzyme flux is represented by the size of the gray arrows. Metabolite concentrations are represented by the size of the circles between enzymes. In this example, increasing the concentration of the second and third enzymes in the pathway increases the titer of the product. Note that decreasing the concentration of intermediate metabolites can be beneficial; this is often the case when intermediates are harmful to the host cell. Increasing enzymes does not always improve product titers and can in fact be detrimental. Figure adapted from (Boyle & Silver 2011).

circuits have also revealed that desired device behavior is highly dependent on the concentration of the device components within cells (Ajo-Franklin et al. 2007; Anderson et al. 2007; Elowitz & Leibler 2000). As a consequence, methods for the control of protein expression levels are essential to metabolic engineering and synthetic biology in general.

Pathway Control: The Rational Approach[4]

Rational pathway tuning approaches attempt to develop a quantitative understanding of a pathway to be engineered and to determine optimal enzyme expression levels *a priori*. In a non-steady-state environment, endogenous metabolic pathways dynamically respond to changes in intracellular metabolite concentrations and changes in the external environment. Forward engineering of such metabolic pathways can be facilitated by a variety of standardized and characterized control elements available to the metabolic engineering community. Regulated gene expression (Beckwith 1967; Jacob & Monod 1961), RNA riboswitches (Mandal & Breaker 2004), and allosteric control of enzyme activity (Monod et al. 1963) provide this control over a wide range of contexts and timescales. Designing similar dynamics into engineered pathways could improve the performance of engineered strains at industrial scales, where reactor conditions are not always uniform (Holtz & Keasling 2010).

For decades, promoter elements have been used to modify gene expression (Reznikoff et al. 1969). In recent years, a number of groups have assembled and characterized promoter libraries for common industrial hosts, such as *Escherichia coli*, *Saccharomyces cerevisiae*, and *Pichia pastoris* (Alper, Fischer, Nevoigt & Stephanopoulos 2005b; Cox et al. 2007; Davis et al. 2010; Hartner et al. 2008; Nevoigt et al. 2006). In each case, native promoters were mutated or recombined to generate a group of promoters of varying strengths. Work has begun to develop standard metrics for promoter characterization, but remains dependent on high-throughput screening of promoter libraries rather than *in silico* prediction (Bayer 2010; Kelly et al. 2009). This issue is compounded by the contextual variability of expression levels in response to environmental factors such as temperature or carbon source (Kelly et al. 2009).

[4] This section is adapted from Boyle and Silver, 2011

Ribosome Binding Sites (RBS) mediate translation initiation, with variation in RBS sequence directly affecting translation efficiency. Thermodynamic models of translation initiation have been generated that now allow *a priori* design of RBS appropriate for a desired expression level. The RBS Calculator (http://salis.psu.edu/software/) generates a customized RBS for a given gene based on the desired translation initiation rate, gene sequence, and host organism. The RBS Calculator was successfully utilized to predict RBS combinations that would permit the desired operation of a synthetic AND gate (Salis et al. 2009), a device that is highly dependent on the expression levels of the inputs to produce AND gate output (Anderson et al. 2007).

Modification of RNA degradation rates can also control steady-state expression levels. In *S. cerevisiae*, the Rnt1p RNAse recognizes and cleaves a specific class of RNA hairpin (Lamontagne et al. 2003). When Rnt1p target hairpins are placed in the untranslated region (UTR) of an mRNA transcript, Rnt1p degradation lowers the effective expression level of the target gene. A library of variable Rnt1p target hairpins has been constructed that permits quantitative control of *S. cerevisiae* gene expression (Babiskin & Smolke 2011).

Modular RNA elements can be designed to provide a dynamic response to intracellular metabolite levels. Riboswitches are natural RNA elements that undergo a conformational change in response to a small-molecule ligand. When riboswitches are part of an mRNA molecule, this conformational change modulates the translation of the mRNA sequence. (Nahvi et al. 2002; Stoddard & Batey 2006). Many riboswitches have been discovered in untranslated regions of mRNAs encoding for metabolic enzymes, offering a post-transcriptional layer of control over enzyme levels.

The potential for RNA-based multisite pathway modulation is exemplified in the 11 known S-adenosylmethionine (SAM) dependent riboswitches of *Bacillus subtilis*. In *B. subtilis*, much of the methionine biosynthesis pathway is regulated by SAM dependent riboswitches. These riboswitches

function primarily through SAM-dependent conformational changes that trigger premature transcriptional termination, although a smaller subset disrupts translation initiation instead. Remarkably, each of the 11 riboswitches is independently tuned to a different SAM concentration. Furthermore, the termination efficiency of each SAM riboswitch in both the ligand bound and unbound conformations are different for each gene (Tomsic et al. 2008). Augmenting engineered metabolic pathways with small-molecule responsive RNA regulators could offer similarly distributed control (Beisel & Smolke 2009).

A variety of synthetic RNA regulators have been designed to control gene expression. Synthetic RNA regulators can interact in *cis* with mRNA via aptamer domains to respond to small molecules (Bayer & Smolke 2005; Win & Smolke 2008), or make use of *trans*-acting RNA elements expressed off of an inducible promoter (Callura et al. 2010; Isaacs et al. 2004). Robust methods have been developed for the selection of RNA aptamer domains (Gilbert & Batey 2005), and modular RNA elements can be combined to generate higher-order behaviors. For example, pairs of RNA aptamer domains alternately promoting or inhibiting translation of a transcript can serve as "bandpass filters," permitting mRNA translation between the range of concentrations set by the aptamer domains (Win & Smolke 2008). Combining promoters and RBS tuned for steady-state performance with dynamically regulated RNA regulators may improve the robustness of engineered pathways.

Pathway Control: The Rationally Irrational Approach[5]

Rational efforts to control pathways with standardized parts have made progress towards making biology "engineerable," yet the complexity of biological systems has kept this process firmly in the trial and error stage. Even synthetic devices with well-defined parameters for desired behavior

[5] This section is adapted from Boyle and Silver, 2011

require exhaustive characterization of the biological components to achieve functionality (Ajo-Franklin et al. 2007; Anderson et al. 2007). However, trial and error through the process of evolution has generated the biological diversity that synthetic biologists seek to redesign. In addition to traditional engineering principles, engineers of biological systems have access to the tools of selection and evolution, which can be leveraged to discover improvements to metabolic pathways. The ability to use these "rationally irrational" approaches is a core advantage to engineering biological systems.

Early metabolic engineering efforts relied on genomic mutagenesis to generate strains with desired properties (Stephanopoulos 1999). If the phenotype of interest is accessible via a single mutation, mutagenesis is an acceptable approach. If the desired phenotype requires multiple mutations, however, the combinatorial expansion of the library size required to identify that phenotype makes untargeted mutagenesis practically infeasible (Dietrich et al. 2010). Generating variation in a targeted subset of the genome enriches the resulting library for mutants with relevant phenotypes (Carr & Church 2009).

Mutagenesis of the cellular transcriptional machinery can be used to adjust gene expression levels. In engineered cells, endogenous regulation often interferes with the functioning of heterologous pathways. Global Transcription Machinery Engineering (gTME) is an approach that modifies relative transcription rates across all genes simultaneously by selectively mutagenizing genes involved in the initiation of transcription. For example, mutagenesis of the *S. cerevisiae* TATA-binding protein *SPT15* and selection for improved ethanol tolerance yielded a mutant with a 20% higher biomass yield than the parent strain (Alper et al. 2006). GTME in *E. coli*, targeting the primary sigma factor σ^{70}, saw similar gains when applied to ethanol tolerance as well as 50% gains when applied to lycopene production (Alper & Stephanopoulos 2007).

A more targeted approach to pathway adjustment is to selectively alter the regulation of pathway genes. Introducing RNAse cleavage sites or hairpin structures that alter mRNA stability into intergenic regions can result in different translation rates for two ORFs on the same mRNA (Smolke et al. 2000). Tunable Intergenic Regions (TIGR) are synthetic RNA constructs that include two hairpins joined by an RNAse cleavage site, and can be used to connect co-transcribed ORFs. Libraries of TIGR elements with a wide variety hairpin structures can be inserted between two co-transcribed genes to screen for optimal translation ratios (Pfleger et al. 2006).

New mutagenesis strategies are enabling iterative and simultaneous mutation of gene regulatory elements. Multiplex Automated Genome Engineering (MAGE) is a high throughput technique for the directed evolution of microbial genomes (Wang et al. 2009). MAGE combines both rational design and directed evolution approaches; specific genomic targets are selected for mutagenesis. For each genomic target, pools of degenerate oligonucleotides that retain homology to the target sequence are electroporated into the cells to be engineered. Multiple pools of oligonucleotides can be combined in a single electroporation step, allowing multiple genomic loci to be modified simultaneously. Iterative rounds of electroporation and growth generate a mixed population of cells with a variety of mutations at loci of interest.

Combinatorial approaches are powerful tools for pathway optimization because they can adjust multiple gene levels simultaneously. Iterative pathway improvement, in which a single gene level is adjusted at a time, can fail to identify global maxima accessible by simultaneous perturbation (Alper & Stephanopoulos 2007). Both gene knockout and up-regulation studies have shown that mutations often interact in a cooperative and non-linear manner with regards to metabolite production (Kennedy et al. 2009). As a further complication, many modern metabolic engineering efforts involve the heterologous expression of enzymes from several different species in an un-optimized host (Agapakis et al. 2010; Bayer et al. 2009; Martin et al. 2003; Ro et al. 2006).

Engineering these chimeric pathways to interface with host metabolism demands that the many factors be adjusted simultaneously.

Generating genomic or pathway-specific variation in gene regulation is only the first step in pathway optimization. Each approach outlined in this section was paired with a screening or selection strategy to identify improved product yields. Pathways that are not observable via high-throughput assays are less amenable to screening approaches. Selection strategies that connect pathway output to cell viability are designed ad hoc, and success is not guaranteed (Dietrich et al. 2010). The lack of generalized methods for pathway screening and selection currently limits the broad application of combinatorial pathway optimization methods.

MODELING METABOLIC NETWORKS[6]

Computational models of metabolism can enhance engineering efforts. The state of the cellular metabolic network is a function of the network topology, the physical properties of enzymes, and the regulation of enzyme levels and activity. While no model has captured the complete complexity of a genome-scale metabolic network, constraint-based approaches have yielded models sufficient for making predictions that can then be validated by rational or irrational experimental approaches.

At the network scale, constraint-based metabolic models utilize the stoichiometry of the metabolome to predict metabolic fluxes. Most constraint-based metabolic models are based on the framework of Flux Balance Analysis (FBA), a technique that simulates the entire metabolic network of an organism (Varma & B. O. Palsson 1994). The only required parameter for an FBA model is a stoichiometric matrix that contains all known metabolic reactions of an organism. Constraints are placed on certain fluxes, defining nutrient availability and relative uptake rates, as well as

[6] This section is adapted from Boyle and Silver, 2009

thermodynamic constraints on the reversibility of reactions. It is assumed that at steady state, the net flux of the system is fixed. The model is then solved for the optimization of an objective function such as maximization of biomass. Since FBA models do not consider enzymatic parameters beyond the stoichiometry of each reaction, the availability of comprehensive databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.com) has fostered the development of FBA models for many organisms (Varma & B. O. Palsson 1994; Duarte et al. 2004; Becker & B. Ø. Palsson 2005; Feist et al. 2007; Lee et al. 2008; Senger & Papoutsakis 2008). Due to the genome-scale power of constraint-based models, *in silico* screens have been applied to predict gene essentiality (Edwards & B. O. Palsson 2000; Thiele et al. 2005; Samal et al. 2006; Becker & B. O. Palsson 2008), and to the related metabolic engineering problem of predicting gene knockouts for strain optimization (Burgard et al. 2003; Alper, Jin, Moxley & Stephanopoulos 2005a; Kennedy et al. 2009).

The successful application of FBA in a variety of organisms demonstrates the utility of constraint-based models in the context of metabolism. However, even in the unlikely case that enzyme kinetics are unimportant to determining metabolic flux, traditional FBA models assume that a cell's entire complement of enzymes are available at all times. Regulation can be modeled implicitly, via methods such as minimization of metabolic adjustment (MOMA), which assumes that regulation will force mutant flux distributions to be as similar to the wild-type distribution as possible (Segrè et al. 2002). Models such as regulatory FBA (rFBA) attempt to explicitly model regulation by switching fluxes on and off based on experimental data of enzyme expression in various growth conditions (Covert et al. 2001; Covert & B. Ø. Palsson 2002; Herrgård et al. 2006).

For the foreseeable future, network scale models of metabolism will serve to augment—but not replace—trial and error design. More complex kinetic models of metabolism suffer from combinatorial increases in complexity when scaled up, making them inadequate for network scale

models (Schuster et al. 1999). In the absence of perfect models, experimental data can be used to reveal key oversights in model predictions.

MEASURING METABOLIC NETWORKS[7]

In many ways, metabolites are the "dark matter" of the cell—their existence, intracellular concentrations, and fluxes are difficult to derive from genomic information and difficult to experimentally measure (Blow 2008). Combining tools for characterizing the status of the metabolome with robust metabolic models is a foundational mission for systems and synthetic biology.

Multiple studies have attempted to reconcile transcriptome data with proteomic or metabolomic measurements in engineered cells (Bradley et al. 2009; Fendt et al. 2010; Ishii et al. 2007; Moxley et al. 2009). The integration of –omics level data with network-scale metabolic models enhances both *a priori* prediction and *post hoc* evaluation of metabolic engineering. In particular, developing quantitative models for the relationship between transcript levels and metabolite pools and fluxes would allow metabolomic data to be inferred from the vast number of microarray datasets that are already available (Yizhak et al. 2010). The chemically uniform nature of mRNA transcripts allows the reliable collection of total mRNA in a single extraction condition. Due to the chemical diversity of small molecules within cells, extraction conditions limit the extent of the metabolome that is observed (Yanes et al. 2011).

Coordination between gene expression and metabolite concentrations appears to be dependent on how is a system is perturbed. Comparison of the *E. coli* transcriptome and metabolome over a range of growth rates revealed that enzyme transcript and protein levels increased with increasing growth rates, while metabolite pools remained steady (Ishii et al. 2007). It

[7] This section is adapted from Boyle and Silver, 2011

was also noted that gene deletions that reverse the flux direction of the pentose phosphate pathway did not significantly alter enzyme levels or metabolite pool sizes. Furthermore, metabolic enzymes did not appreciably up-regulate to compensate for enzyme knockouts (Ishii et al. 2007). On the other hand, a study comparing the transcriptome and metabolome of *S. cerevisiae* during carbon and nitrogen starvation observed coordinated changes in expression and metabolite levels(Bradley et al. 2009).

Further work is required to identify contexts in which transcript levels correlate to metabolite concentrations. Comparison of the above studies suggests a differential metabolic response to enzyme knockouts versus shifting media conditions (Bradley et al. 2009; Ishii et al. 2007); this could be a consequence of evolutionary selection for robustness against condition changes (Cornelius et al. 2011; Segrè et al. 2002). Alternatively, *E. coli* and *S. cerevisiae* may simply respond differently to metabolic perturbations.

A significant confounding issue is that major metabolic flux alterations can occur without major shifts in enzyme or metabolite concentrations (Fell 1997; Ishii et al. 2007). Two recent studies measured transcriptomic and metabolomic shifts in *S. cerevisiae* in response to the deletion of global regulatory genes rather than enzymes (Fendt et al. 2010; Moxley et al. 2009). Following the deletion of the Gcn4p, a global stress response regulator, metabolites that were involved in many enzymatic reactions had the most influence on flux through their respective pathways (Moxley et al. 2009). This raises the possibility of utilizing network topology to inform metabolic engineering. In the case of central carbon metabolism, deletion of the glycolysis-activating transcription factor Gcr2p showed a negative correlation between enzyme levels and associated metabolite levels (Fendt et al. 2010). This could be indicative of a buffering phenomenon, in which changes in metabolite pools counteract enzyme concentration changes to maintain a steady pathway flux.

A grand unifying theory of metabolism has not yet arisen from these meta-omics studies. It is possible that a truly general relationship between gene expression and metabolic concentrations does not exist. Overall, however, it appears that evolved metabolic networks are quite robust in response to genetic and environmental perturbations. This is corroborated by many of the metabolic engineering efforts that we have reviewed, in which multiple perturbations were required to improve product yields. Integrative data from –omics scale datasets may help to identify genes that contribute to the observed resistance to perturbations in pathways of interest.

TOWARDS AN ENGINEERING DESIGN CYCLE FOR BIOLOGY

The work presented in this dissertation is intended to assist the ongoing effort to develop true engineering design cycles for biological systems. Predictable rational design is a universal feature of all engineering disciplines. Rational design of biological systems continues to improve as new models and genomic engineering techniques are developed. Irrational design, based on mutagenesis and directed evolution, is an advantageous quality of biological systems that should be used to the engineer's benefit. This dissertation presents applications of both rational and irrational design in a variety of contexts.

During my time in the Silver laboratory we established ourselves as a "species independent" synthetic biology lab: we chose to work with organisms based on genetic tractability, chassis-specific metabolic properties, and applicability to an interesting biological problem. As a result, my fellow labmates and I have worked on a wide array of organisms over the last six years. In particular, I have personally worked on *S. cerevisiae*, *E. coli*, *Synechococcus elongatus*, and *Arabidopsis thaliana* as a graduate student (Figure 1.5). We extracted and cloned DNA from *acetobutylicum*[8] to *Zea mays* to obtain parts

[8] *Clostridum acetobutylicum*, to be exact.

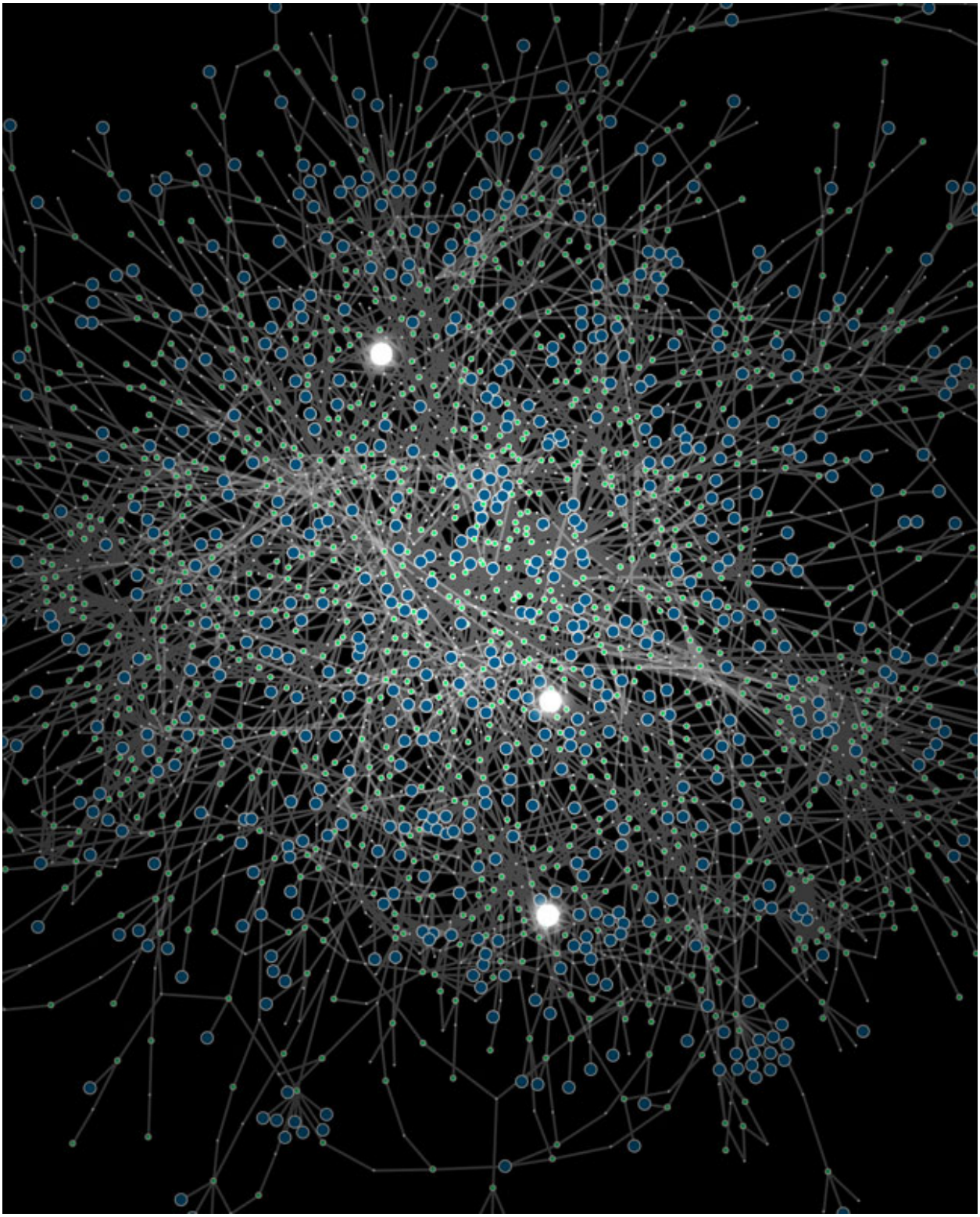


Figure 1.3 The Yeast Metabolic Network. Blue nodes represent genes, green nodes represent enzymatic reactions. The three white nodes represent knockouts selected by the FBA model used in Chapter 2 to improve formic acid production. This image was originally published as cover art for *Genetics* **183**(1) (2009).

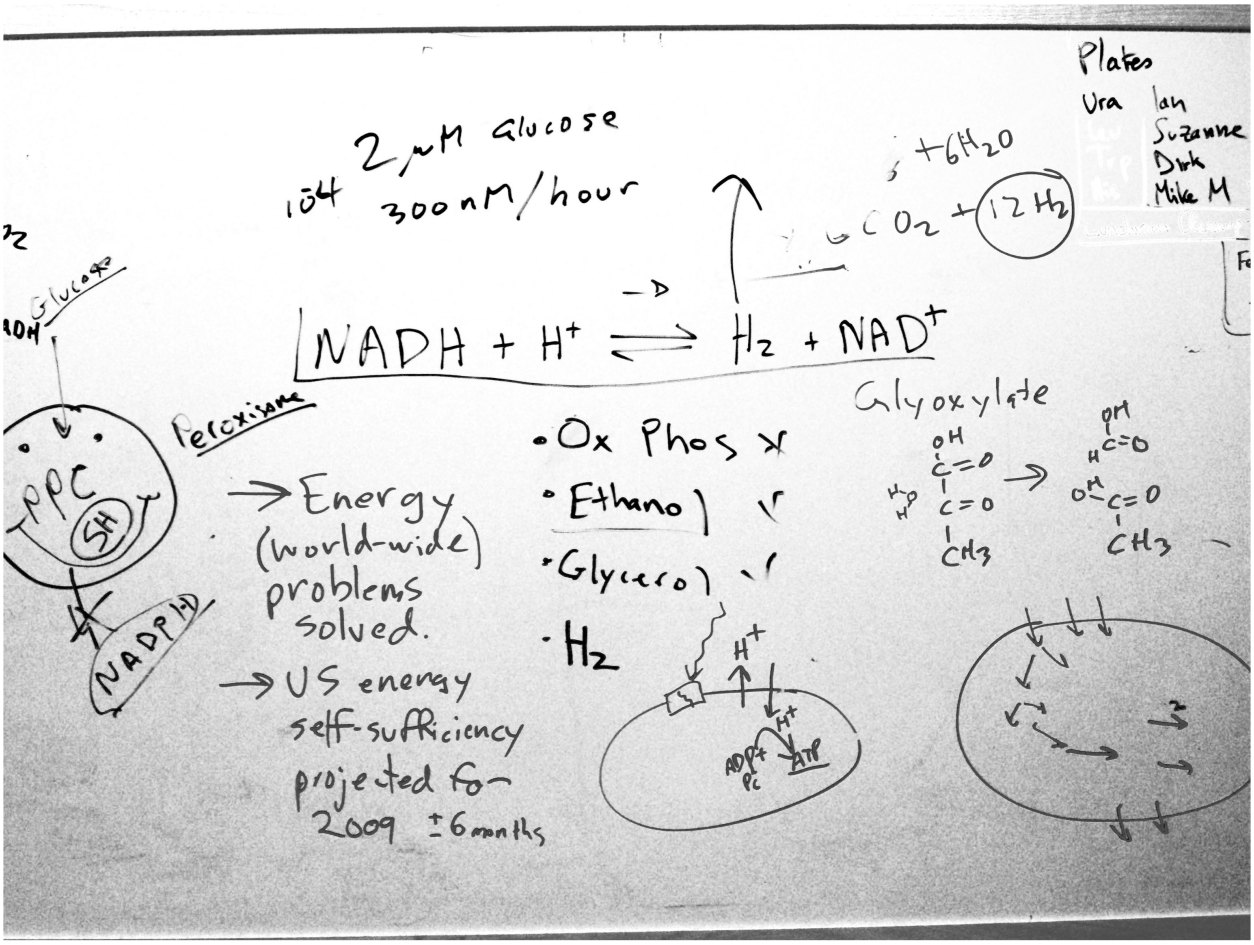


Figure 1.4 The Silver Lab Metabolic Engineering Master Plan (2007)

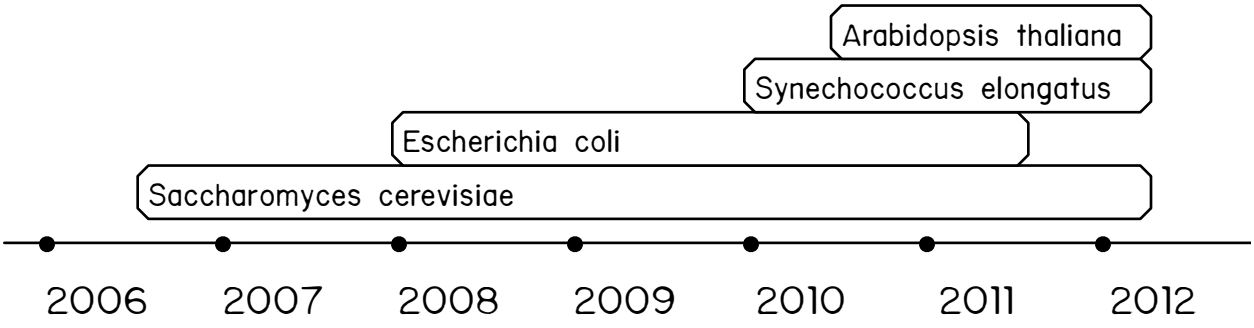


Figure 1.5 Approximate time period in which each organism was utilized as a chassis.

to test in these chasses. We are entering an era of "Synthetic Metagenomics," in which novel metabolic pathways can be constructed from optimal combinations of enzymes that originated in many different species (Bayer et al. 2009). Studying and engineering metabolism in diverse contexts also highlights the features of metabolic networks that are common to all species.

In chapter 2, I present a project in which a computational model of *S. cerevisiae* metabolism was used to identify gene deletions that would lead to increased formic acid production. Model predictions were validated experimentally, demonstrating that the top hit in our computational screen did in fact improve formic acid titers. This work demonstrates the utility of network-scale approaches in metabolic engineering: a genome-scale FBA model of metabolism was used to establish our predictions, and microarray analysis identified unanticipated regulatory interventions that negatively impacted product yields.

In chapter 3, I present my contributions to our laboratory's efforts to optimize the function of [FeFe]-hydrogenases. We tested a variety of synthetic biology concepts during the course of this work: we cloned enzymes from numerous species to optimize the heterologous production of hydrogen in *E. coli*, systematically deleted host metabolic reactions that competed with our synthetic hydrogen pathway, and designed genetic selections to "irrationally" identify hydrogenase mutants with improved oxygen tolerance.

In chapter 4, I demonstrate the application of network-scale measurements to understanding metabolic networks. Targeted liquid chromatography/mass spectrometry was used to track over 200 metabolites in the model cyanobacterium *S. elongatus* PCC 7942. While this work has metabolic engineering implications in that it improves our understanding of *S. elongatus* metabolism, it also addresses basic questions regarding metabolic responses to transcriptional regulation.

I have also included several papers as Appendices that I wrote during the course of my graduate work that are supplemental to the main narrative of this dissertation. Appendices B and C

include the full text of two reviews on synthetic biology that contributed to this introduction chapter. Appendix D presents a review that I wrote with Devin Burrill on the application of Synthetic Biology to human metabolic disease. Appendix E presents a research paper that I wrote with the members of the 2010 Harvard team for the International Genetically Engineered Machines (iGEM) competition. In Appendix F, Christina Agapakis and I review spatial methods for optimizing metabolic pathways.

REFERENCES

Agapakis, C.M. & Silver, P.A., 2009. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems*, 5(7), pp.704–713.

Agapakis, C.M. et al., 2010. Insulation of a synthetic hydrogen metabolism circuit in bacteria. *Journal of Biological Engineering*, 4, p.3.

Ajo-Franklin, C.M. et al., 2007. Rational design of memory in eukaryotic cells. *Genes & Development*, 21(18), pp.2271–2276.

Alper, H. & Stephanopoulos, G., 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metabolic Engineering*, 9(3), pp.258–267.

Alper, H. et al., 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science (New York, NY)*, 314(5805), pp.1565–1568.

Alper, H., Jin, Y.-S., Moxley, J.F. & Stephanopoulos, G., 2005a. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metabolic Engineering*, 7(3), pp.155–164.

Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G., 2005b. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 102(36), pp.12678–12683.

Alper, H., Miyaoku, K. & Stephanopoulos, G., 2005c. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nature Biotechnology*, 23(5), pp.612–616.

Anderson, J.C., Voigt, C.A. & Arkin, A.P., 2007. Environmental signal integration by a modular AND gate. *Molecular Systems Biology*, 3, p.133.

Andrianantoandro, E. et al., 2006. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2, p.2006.0028.

Babiskin, A.H. & Smolke, C.D., 2011. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Molecular Systems Biology*, 7, p.471.

Bayer, T.S., 2010. Grand Challenge Commentary: Transforming biosynthesis into an information science. *Nature Chemical Biology*, 6(12), pp.859–861.

Bayer, T.S. & Smolke, C.D., 2005. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology*, 23(3), pp.337–343.

Bayer, T.S. et al., 2009. Synthesis of methyl halides from biomass using engineered microbes. *Journal of the American Chemical Society*, 131(18), pp.6508–6515.

Becker, S.A. & Palsson, B.O., 2008. Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Systems Biology*, 2, p.14.

Becker, S.A. & Palsson, B.Ø., 2005. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiology*, 5, p.8.

Beckwith, J.R., 1967. Regulation of the lac operon. Recent studies on the regulation of lactose metabolism in Escherichia coli support the operon model. *Science (New York, NY)*, 156(3775), pp.597–604.

Beisel, C.L. & Smolke, C.D., 2009. Design principles for riboswitch function. *PLoS Computational Biology*, 5(4), p.e1000363.

Blow, N., 2008. Metabolomics: Biochemistry's new look. *Nature*, 455(7213), pp.697–700.

Boyle, P.M. & Silver, P.A., 2009. Harnessing nature's toolbox: regulatory elements for synthetic biology. *Journal of the Royal Society, Interface / the Royal Society*, 6 Suppl 4, pp.S535–46.

Boyle, P.M. & Silver, P.A., 2011. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metabolic Engineering*.

Bradley, P.H. et al., 2009. Coordinated concentration changes of transcripts and metabolites in Saccharomyces cerevisiae. *PLoS Computational Biology*, 5(1), p.e1000270.

Burgard, A.P., Pharkya, P. & Maranas, C.D., 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6), pp.647–657.

Burrill, D.R., Boyle, P.M. & Silver, P.A., 2011. A New Approach to an Old Problem: Synthetic Biology Tools for Human Disease and Metabolism. *Cold Spring Harbor symposia on quantitative biology*.

Callura, J.M. et al., 2010. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), pp.15898–15903.

Carlson, R.H., 2010. *Biology is Technology: The Promise, Peril, and New Business of*

Engineering Life., Cambridge, MA: Harvard University Press.

Carothers, J.M., Goler, J.A. & Keasling, J.D., 2009. Chemical synthesis using synthetic biology. *Current Opinion in Biotechnology*, 20(4), pp.498–503.

Carr, P.A. & Church, G.M., 2009. Genome engineering. *Nature Biotechnology*, 27(12), pp.1151–1162.

Cornelius, S.P., Lee, J.S. & Motter, A.E., 2011. Dispensability of Escherichia coli's latent pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), pp.3124–3129.

Covert, M.W. & Palsson, B.Ø., 2002. Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *The Journal of biological chemistry*, 277(31), pp.28058–28064.

Covert, M.W., Schilling, C.H. & Palsson, B., 2001. Regulation of gene expression in flux balance models of metabolism. *Journal of theoretical biology*, 213(1), pp.73–88.

Cox, R.S., Surette, M.G. & Elowitz, M.B., 2007. Programming gene expression with combinatorial promoters. *Molecular Systems Biology*, 3, p.145.

Davis, J.H., Rubin, A.J. & Sauer, R.T., 2010. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research*.

Dekel, E. & Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050), pp.588–592.

Dietrich, J.A., Mckee, A.E. & Keasling, J.D., 2010. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annual review of biochemistry*, 79, pp.563–590.

Drubin, D.A., Way, J.C. & Silver, P.A., 2007. Designing biological systems. *Genes & Development*, 21(3), pp.242–254.

Duarte, N.C., Herrgård, M.J. & Palsson, B.Ø., 2004. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome research*, 14(7), pp.1298–1309.

Edwards, J.S. & Palsson, B.O., 2000. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*, 1, p.1.

Elowitz, M.B. & Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), pp.335–338.

Endy, D., 2005. Foundations for engineering biology. *Nature*, 438(7067), pp.449–453.

Feist, A.M. et al., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3, p.121.

Fell, D., 1997. *Understanding the Control of Metabolism* K. Snell, ed. London: Portland Press.

Fendt, S.-M. et al., 2010. Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Molecular Systems Biology*, 6, p.356.

Forster, A.C. & Church, G.M., 2006. Towards synthesis of a minimal cell. *Molecular Systems Biology*, 2, p.45.

Gilbert, S.D. & Batey, R.T., 2005. Riboswitches: natural SELEXion. *Cellular and molecular life sciences : CMLS*, 62(21), pp.2401–2404.

Hartner, F.S. et al., 2008. Promoter library designed for fine-tuned gene expression in Pichia pastoris. *Nucleic Acids Research*, 36(12), p.e76.

Haynes, K.A. & Silver, P.A., 2009. Eukaryotic systems broaden the scope of synthetic biology. *The Journal of Cell Biology*, 187(5), pp.589–596.

Herrgård, M.J. et al., 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome research*, 16(5), pp.627–635.

Holtz, W.J. & Keasling, J.D., 2010. Engineering static and dynamic control of synthetic pathways. *Cell*, 140(1), pp.19–23.

Isaacs, F.J. et al., 2004. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7), pp.841–847.

Ishii, N. et al., 2007. Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science (New York, NY)*, 316(5824), pp.593–597.

Jacob, F. & Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3, pp.318–356.

Jones, K.L., Kim, S.W. & Keasling, J.D., 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metabolic Engineering*, 2(4), pp.328–338.

Keasling, J.D., 2010. Manufacturing molecules through metabolic engineering. *Science (New York, NY)*, 330(6009), pp.1355–1358.

Keasling, J.D., 2008. Synthetic biology for synthetic chemistry. *ACS chemical biology*, 3(1), pp.64–76.

Kelly, J.R. et al., 2009. Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering*, 3, p.4.

Kennedy, C.J. et al., 2009. Systems-level engineering of nonfermentative metabolism in yeast. *Genetics*, 183(1), pp.385–397.

Lamontagne, B. et al., 2003. Sequence dependence of substrate recognition and cleavage by yeast RNase III. *Journal of Molecular Biology*, 327(5), pp.985–1000.

Lee, J. et al., 2008. Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. *Applied Microbiology and Biotechnology*, 80(5), pp.849–862.

Mandal, M. & Breaker, R.R., 2004. Gene regulation by riboswitches. *Nature reviews Molecular cell biology*, 5(6), pp.451–463.

Martin, V.J.J. et al., 2003. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology*, 21(7), pp.796–802.

Monod, J., Changeux, J.P. & Jacob, F., 1963. Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 6, pp.306–329.

Moxley, J.F. et al., 2009. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16), pp.6477–6482.

Nahvi, A. et al., 2002. Genetic control by a metabolite binding mRNA. *Chemistry & Biology*, 9(9), p.1043.

Nevoigt, E. et al., 2006. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, 72(8), pp.5266–5273.

Pfleger, B.F. et al., 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology*, 24(8), pp.1027–1032.

Qin, J. et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), pp.59–65.

Raab, R.M., Tyo, K. & Stephanopoulos, G., 2005. Metabolic engineering. *Advances in biochemical engineering/biotechnology*, 100, pp.1–17.

Reznikoff, W.S. et al., 1969. A mechanism for repressor action. *Journal of Molecular Biology*, 43(1), pp.201–213.

Ro, D.-K. et al., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086), p.940.

Salis, H.M., Mirsky, E.A. & Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10), pp.946–950.

Samal, A. et al., 2006. Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinformatics*, 7, p.118.

Schuster, S., Dandekar, T. & Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology*, 17(2), pp.53–60.

Segrè, D., Vitkup, D. & Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), pp.15112–15117.

Senger, R. & Papoutsakis, E., 2008. Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis. *Biotechnology and Bioengineering*.

Smolke, C.D., Carrier, T.A. & Keasling, J.D., 2000. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Applied and Environmental Microbiology*, 66(12), pp.5399–5405.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. *Metabolic Engineering*, 1(1), pp.1–11.

Stoddard, C.D. & Batey, R.T., 2006. Mix-and-match riboswitches. *ACS chemical biology*, 1(12), pp.751–754.

Thiele, I. et al., 2005. Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *Journal of Bacteriology*, 187(16), pp.5818–5830.

Tomsic, J. et al., 2008. Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. *Journal of Bacteriology*, 190(3), pp.823–833.

Varma, A. & Palsson, B.O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology*, 60(10), pp.3724–3731.

Wang, H.H. et al., 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257), pp.894–898.

Win, M.N. & Smolke, C.D., 2008. Higher-order cellular information processing with synthetic RNA devices. *Science (New York, NY)*, 322(5900), pp.456–460.

Yadav, V.G. & Stephanopoulos, G., 2010. Reevaluating synthesis by biology. *Current Opinion in Microbiology*.

Yanes, O. et al., 2011. Expanding Coverage of the Metabolome for Global Metabolite Profiling. *Analytical chemistry*.

Yizhak, K. et al., 2010. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics (Oxford, England)*, 26(12), pp.i255–60.

Zaslaver, A. et al., 2004. Just-in-time transcription program in metabolic pathways. *Nature Genetics*, 36(5), pp.486–491.

Chapter 2

Systems-level engineering of non-fermentative metabolism in yeast[1]

ABSTRACT

We designed and experimentally validated an *in silico* gene deletion strategy for engineering endogenous one-carbon (C1) metabolism in yeast. We used constraint-based metabolic modeling and computer-aided gene knockout simulations to identify five genes (*ALT2*, *FDH1*, *FDH2*, *FUM1*, and *ZWF1*), which, when deleted in combination, predicted formic acid secretion in *Saccharomyces cerevisiae* under aerobic growth conditions. Once constructed, the quintuple mutant strain showed the predicted increase in formic acid secretion relative to a formate dehydrogenase mutant (*fdh1 fdh2*), while formic acid secretion in wild-type yeast was undetectable. Gene expression and physiological data generated post hoc identified a retrograde response to mitochondrial deficiency, which was confirmed by showing Rtg1-dependent NADH accumulation in the engineered yeast strain. Formal pathway analysis combined with gene expression data suggested specific modes of regulation that govern C1 metabolic flux in yeast. Specifically, we identified coordinated transcription regulation of C1 pathway enzymes and a positive flux control coefficient for the branchpoint enzyme 3-phosphoglycerate dehydrogenase (PGDH). Together, these results demonstrate that constraint-

[1] This chapter was originally published as Kennedy, C. J.*, Boyle, P. M.*, Waks, Z. & Silver, P. A. Systems-level engineering of nonfermentative metabolism in yeast. Genetics 183, 385–397 (2009).
*Equal contribution.
Contributions: CJK, PMB, and PAS designed experiments and wrote the paper. CJK and PMB did the computational work jointly. CJK, PMB, and ZW performed experiments.

based models can identify seemingly unrelated mutations, which interact at a systems level across subcellular compartments to modulate flux through non-fermentative metabolic pathways.

INTRODUCTION

Formic acid and one-carbon metabolism: Formic acid is an important intracellular metabolite that has been adapted for specific functions in different organisms. It is produced and secreted in small amounts as a fermentation byproduct by bacteria in the family *Enterobacteriaceae* (Leonhartsberger et al., 2002) and in large quantities as an irritant and pheromone by ants (Hefetz and Blum, 1978). Formic acid is used commercially as a preservative in animal feed and has a potential use as a precursor to hydrogen, since it is one of only a few biological molecules with sufficient reducing potential (Milliken and May, 2007). The main pathway for biohydrogen production during mixed acid fermentation in *E. coli* proceeds through a formic acid intermediate: a product of the reaction catalyzed by pyruvate formate lyase (PFL, EC 2.3.1.54) (Birkmann et al., 1987).

As yeast (and other eukaryotes) lack a PFL homologue, their primary source of formic acid is through tetrahydrofolate (THF)-mediated one-carbon (C1) reactions present in the mitochondria (McNeil et al., 1996). In mammalian cells C1 metabolism is responsible for up to 90% of single carbon units required for nucleotide biosynthesis (Fu et al., 2001). The first reaction in this pathway is catalyzed by the branchpoint enzyme 3-phosphoglycerate dehydrogenase (PGDH, EC 1.1.1.95) encoded by the yeast isozymes *SER3* and *SER33*. The NAD-dependent oxidation reaction catalyzed by PGDH is non-fermentative: oxygen, rather than organic substrate, acts as the final electron acceptor to maintain redox homeostasis under conditions where high levels of serine and formic acid are synthesized from the glycolytic intermediate 3-phosphoglycerate (3PG; Peters-Wendisch et al., 2005).

Constraint-based modeling: Constraint-based (stoichiometric) models are capable of describing systems-level properties of metabolic networks without requiring specific information about molecular mechanism or reaction-specific kinetics. As many of these parameters are often unknown, constraint-based methods have advantages over their kinetic counterparts as practical tools for developing systems-level metabolic engineering strategies. These models rely on well-annotated genomic sequences in order to define sets of metabolites and the stoichiometric matrix of known biochemical reactions. Once these are defined, feasible assumptions about quasi steady-state optimality are all that is necessary to predict reaction rates for the entire system. Combinatorial enzyme deletion phenotypes can be explored systematically by constraining specific enzyme-reaction fluxes to zero (for example Forster et al., 2003; Edwards and Palsson, 2000). This approach provides reasonable approximations of genome-wide biochemical processes in several model organisms (Becker and Palsson, 2008; Motter et al., 2008; Hjersted et al., 2007; Resendis-Antonio et al., 2007; Oh et al., 2007; Bro et al., 2006; Duarte et al., 2004; Edwards and Palsson, 2000).

Constraint-based methods provide a solid mathematical foundation for identifying important properties of biochemical pathways. Under a certain set of stoichiometric constraints, metabolic networks can be decomposed into a finite number of elementary flux modes (EFMs) or extreme pathways (Papin et al., 2002). The properties of EFMs have important biological implications. EFMs are the unique set of non-decomposeable pathway flows for a given biochemical network (Shuster, 1999). In biological terms, EFMs are modular units of pathway function—minimal sets of enzymes required to catalyze whole metabolic reactions. Because metabolic pathways are highly integrated, the number of possible pathways connecting reactants and products grows exponentially. Thus EFM analysis is only computationally tractable for individual pathways or small metabolic subnetworks (Klamt and Stelling, 2002).

Past metabolic engineering efforts in eukaryotic microbes have sought to control flux

through anaerobic pathways for increased production of metabolites produced by fermentation. These include lactate (Ishida et al., 2006; van Maris et al., 2004), malate (Zelle et al., 2008), isoprenoids (Kizer et al., 2008; Herrero et al., 2008; Shiba et al., 2007), glycerol (Cordier et al., 2007; Geertman et al., 2006), and ethanol (Alper et al., 2006 and Bro et al., 2006), among others. Although non-fermentative byproducts represent a class of biologically interesting and commercially attractive small molecules, efforts aimed at engineering microbes for increased production of these metabolites are comparatively infrequent.

Reactions that produce the major one-carbon donors serine, glycine, and formic acid are often duplicated in the cytoplasm and mitochondrion (Christensen and MacKenzie, 2006). Flux through these reactions is generally oxidative in the mitochondria and reductive in the cytoplasm; however, C1 metabolic pathways are under considerable regulatory control and can be adapted to specific genetic backgrounds and growth conditions (Piper et al., 2000; Kastanos et al., 1997). Two groups have independently shown that C1 enzymes are controlled dynamically by glycine at the transcription level. Upon glycine withdrawal many enzymes involved in C1 metabolism are strongly repressed; a regulatory event that requires the transcription factor Bas1 (Subramanian et al., 2005). Under conditions of glycine induction Gelling et al. (2004) noticed a similar pattern of C1-enzyme differential expression, however Bas1 was not required for the observed effect in this case. In both of these studies the intact cytosolic serine hydroxymethyltransferase Shm2 (EC 2.1.2.1) was necessary for glycine-dependent changes in C1 enzyme expression (Subramanian et al., 2005; Gelling et al., 2004). Though contradicting evidence exists, results reported thus far demonstrate the dynamic control of C1 metabolism in eukaryotes.

To investigate metabolic engineering strategies for controlling biosynthetic flux through a non-fermentative pathway, we chose to construct strains of *S. cerevisiae* that increase flux through C1 metabolism. We chose a constraint-based modeling approach to develop genetic engineering

strategies leading to increased production of formic acid. We experimentally validated our modeling strategy and identified specific transcription control mechanisms that govern C1 metabolism in the engineered strain.

RESULTS

Pathway analysis of C1 metabolism in yeast: We performed limited pathway analysis of the yeast metabolic network to identify Eq. (1), which represents the complete oxidation of glucose into formic acid (see Methods):

$$C_6H_{12}O_6 + 4O_2 \rightarrow 4HCOOH + 2CO_2 + 2H_2O + \text{energy} \qquad \text{Eq. (1)}$$

Due to the quasi steady-state assumption, individual reaction rates in Eq. (1) are likely to be correlated during C1-mediated formic acid secretion; a condition that strongly suggests coordinated regulation of enzymes in this pathway (Schuster et al., 2002). Co-regulation of enzymes involved in glycolysis is well characterized; however, it is not fully understood if and how endogenous programs coordinate C1 metabolism to affect formic acid biosynthesis.

Theoretically, maximum flux through Eq. (1) would result in 4 formic acid molecules per glucose, a two-fold yield increase compared to PFL-catalyzed reactions associated with mixed acid fermentation in *E. coli* (Birkmann et al., 1987). This may have important biotechnological applications for biofuel production. Assuming 100% conversion of exogenous formic acid into hydrogen, a two-step conversion process using endogenous hydrogenases in *E. coli* could result in 4 H_2 per glucose (Yoshida et al., 2005; Waks and Silver, 2009). Formic acid production in yeast is relatively low and secretion is essentially non-existent (Blank et al., 2005); however, there is reason to believe that engineering C1 metabolism to produce high levels of formic acid is achievable. Various insect species regulate homologous pathways to produce large quantities of formic acid for the purposes of defense and communication (Hefetz and Blum, 1978).

Rank	Genotype	Efficiency (%)[a]
1	*fdh1 fdh2 alt2 fum1 zwf1*	72.3
2	*fdh1 fdh2 aat2 fum1 zwf1*	72.2
3	*fdh1 fdh2 cat2 fum1 zwf1*	72.0
4	*fdh1 fdh2 cat2 fum1 rpe1*	71.7
5	*fdh1 fdh2 cat2 fbp1 fum1*	30.5
6	*fdh1 fdh2 cat2 yat2 slc1*	2.4
7	*fdh1 fdh2 cat2 yat2 cho1*	2.3
8	*fdh1 fdh2 cat2 yat2 alt2*	1.2

Table 2.1 Gene combinations affecting C1 metabolic flux and formic acid secretion identified through *in silico* knockout simulation

[a] 100% efficiency is defined as 4 formic acid molecules per glucose.

Strain	Genotype[a]	Reference
PSY3639	*fdh1(41, 1091)::loxP fdh2(41, 1091)::loxP*	Overkamp et al., 2002
PSY3640	*zwf1::loxP*	this study
PSY3641	*zwf1::loxP fum1::loxP*	this study
PSY3642	*zwf1::loxP fum1::loxP alt2::loxP*	this study
PSY3650	*zwf1::loxP fum1::loxP alt2::loxP rtg1::KanMX*	this study
PSY3653	*fum1::KanMX*	this study

Table 2.2 Yeast strains

[a] All mutations are present in the CEN.PK113-7D MATa URA3 HIS3 LEU2 TRP1 MAL2 SUC2 genetic background. PSY3640-PSY3642, PSY3650 , and PSY3653 were constructed from PSY3639.

A model-driven metabolic engineering strategy to increase endogenous formic acid secretion: To formulate genetic engineering strategies leading to increased production of formic acid in yeast, we used the compartmentalized metabolic model *i*ND750 (Duarte et al., 2004) and an iterative gene knockout simulation strategy to identify combinatorial enzyme deletions predicted to significantly increase formic acid secretion (see Methods for details). As an exhaustive screen through all triple knockout combinations would have been experimentally infeasible, we used flux balance analysis (FBA) to screen combinatorial knockouts *in silico*. We identified several gene knockout combinations, which predicted non-zero secretion rates of formic acid (Table 2.1). In all

cases eliminating the NAD-dependent formate dehydrogenase reaction FDH (EC 1.2.1.2) was required for secretion of formic acid. This is not surprising as FDH is thought to protect yeast from formic acid toxicity by catalyzing its irreversible oxidation to CO_2 (Overkamp et al., 2002).

We chose to proceed by constructing the mutant strain predicted to have the highest formic acid production efficiency (Table 2.1). In addition to *FDH1* and *FDH2*, three other genes were targeted for mutation by serial gene replacement (Table 2.2): *ALT2*, a putative cytoplasmic alanine transaminase (EC 2.6.1.2); *FUM1*, fumarase (EC 4.2.1.2); and *ZWF1*, glucose-6-phosphate dehydrogenase (EC 1.1.1.49). Unlike FDH, these three genes (reactions) function across subcellular compartments at distantly located positions within the metabolic network and are not obviously associated with formic acid biosynthesis or C1 metabolism.

A predicted increase in flux to formic acid is achieved through non-intuitive interactions between *alt2*, *fum1*, and *zwf1* (Figure 2.1). Although the protein encoded by *FUM1* is both cytoplasmic and mitochondrial (Stein et al., 1994), the model predicted several effects specifically related to its mitochondrial function: (i) decoupling the respiratory chain resulting in (ii) decreased flux into the key TCA cycle intermediate alpha-ketoglutarate (AKG) and (iii) increased flux into 3-phosphoglycerate (3PG). Flux through PGDH compensates for loss of *FUM1* by balancing 3PG and generating the cytoplasmic AKG—via phosphoserine transaminase (PST, EC 2.6.1.52) that is necessary for growth. PST catalyzes the transamination of 3-phosphonooxypyruvate using glutamate as a cofactor, which is balanced by eliminating the competitive reaction associated with *ALT2*. Removing *ZWF1* eliminates direct flux to the pentose phosphate pathway thereby increasing flux into 3PG, which is balanced by serine/glycine biosynthesis leading to the complete oxidation of glucose into formic acid, carbon dioxide, and biomass. In the absence of *FDH1* and *FDH2*, intracellular formic acid is balanced by secretion into the media.

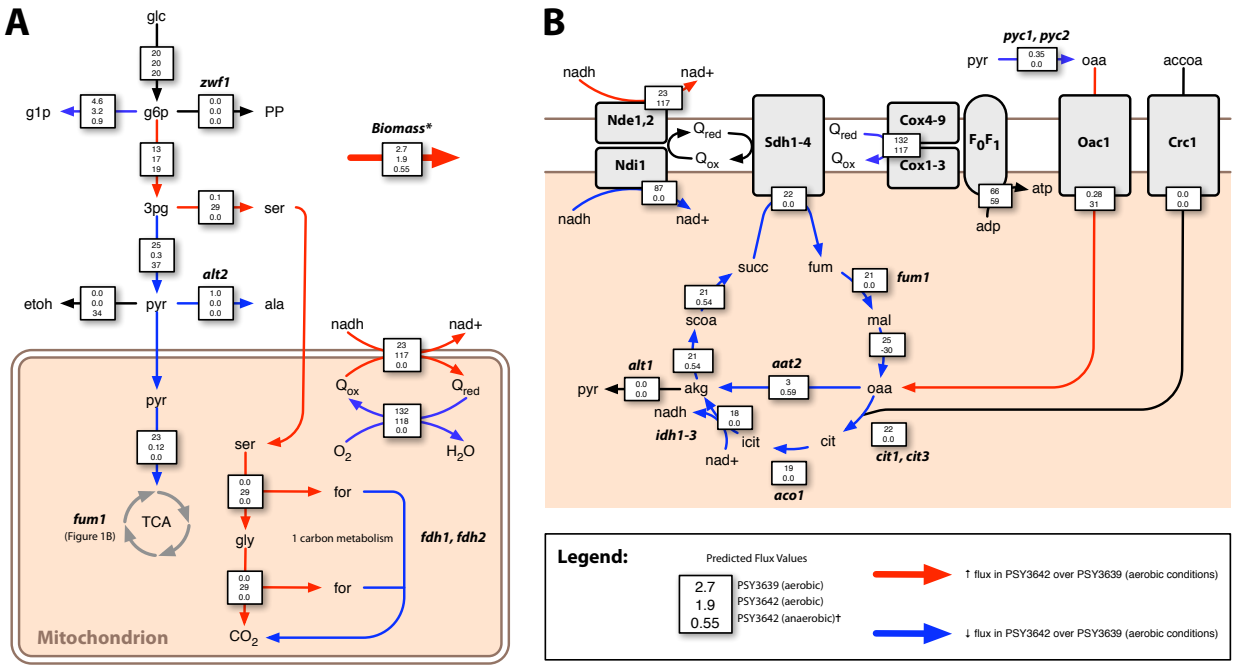


Figure 2.1 Constraint based modeling predicts mutations that redirect flux through serine/glycine biosynthesis and C1 metabolism leading to increased aerobic formic acid secretion. Arrows denote key cytoplasmic (A) and mitochondrial (A and B) reactions for which predicted flux is higher (red) or lower (blue) in PSY3642 compared to PSY3639 (see text for details). Boxes superimposed over arrows contain flux values for three simulated conditions: (i) PSY3639 (aerobic), (ii) PSY3642 (aerobic), and (iii) PSY3642 (anaerobic). Flux values are relative to a constant glucose uptake rate of 20 mmol gDW^{-1} hr^{-1}. *See Supplementary Table S2.2 for a complete list of predicted fluxes and metabolites required for growth (the Biomass equation). †Only given in (A). Abbreviations (not provided in the text): glc, glucose; g1p, glucose-1-phosphate; g6p, glucose-6-phosphate; ser, serine; etoh, ethanol; pyr, pyruvate; ala, alanine; gly, glycine; for, formate; nad, nicotinamide adenine dinucleotide; Q, quinone; accoa, acetyl-CoA; cit, citrate; isocit, isocitrate; succ, succinate; fum, fumarate; and mal, malate.

Consistent with Eq. (1), the FBA model predicts that formic acid secretion is oxygen-dependent. Excess reducing equivalents in the form of 8 cytoplasmic NADH are balanced aerobically rather than using organic substrate as a final electron acceptor. Accordingly, the model predicts increased flux through reactions catalyzed by the external mitochondrial NADH dehydrogenases Nde1 and Nde2 as well as downstream electron transport chain components.

Model validation confirms that elevated formic acid secretion requires aerobic respiration: Results comparing formic acid secretion in PSY3639 and PSY3642 revealed broad qualitative

agreement with two important model predictions: (i) mutations in *alt2, fum1* and *zwf1* appear to interact in a combinatorial manner to enhance formic acid production and (ii) this enhancement was oxygen-dependent.

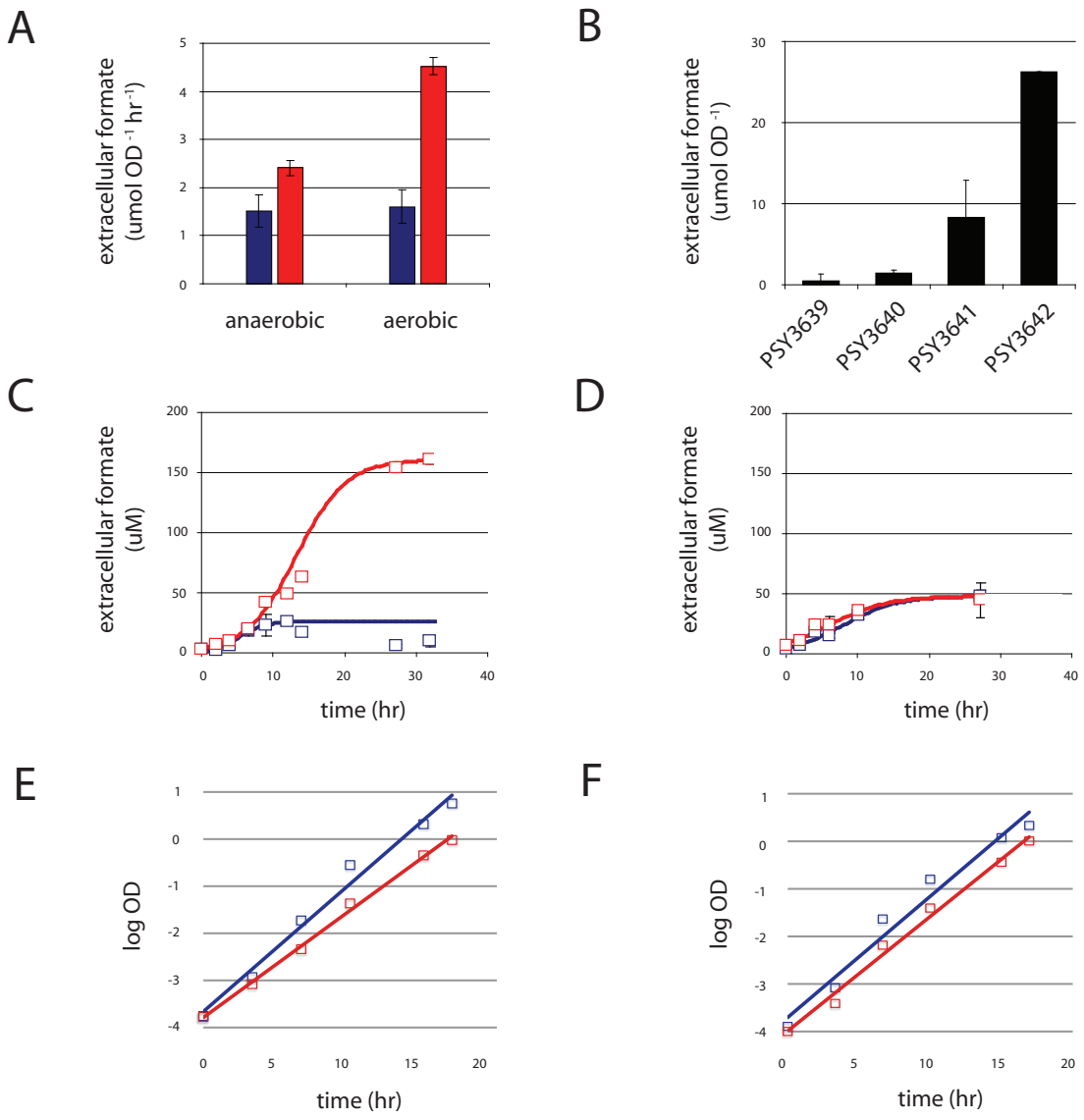


Figure 2.2 Experimental validation of model predictions. Given are formic acid secretion rates under aerobic and anaerobic growth conditions (A), total extracellular formic acid production for several strains (B) including PSY3639 (blue) and PSY3642 (red) grown aerobically (C) and anaerobically (D). Growth rates are given for PSY3639 and PSY3642 grown aerobically (E) and anaerobically (F). Genotypes for strains in B are listed in Table 2.2. Fit-curves in C-F were calculated using logistic regression. Data represent the average of three biological replicates + SD. A paired two-tailed t-test was used to test for statistical significance in (A).

The rate of formic acid secretion measured during log-phase growth was significantly (3-fold) higher in PSY3642 compared to PSY3639 ($p < 0.01$) and this change was dependent on aerobic growth conditions (Figure 2.2A). Furthermore, formic acid secretion increased non-linearly with enzyme loss indicating a cumulative increase in formic acid production that resulted from eliminating combinatorial interactions between the deleted enzymes (Figure 2.2B). Because FBA models flux at quasi steady-state, derived predictions are generally applicable only during log-phase growth. However, comparisons of PSY3642 and PSY3639 made after saturation (> 30 hours of continuous growth) revealed a striking 16.5-fold increase in extracellular formic acid concentration (Figure 2.2C). Consistent with model predictions, this difference in total formic acid secretion was only observed under aerobic growth conditions (Figure 2.2D).

Gene	ORF	Function	Fold-Change
ADE4	YMR300C	Phosphoribosylpyrophosphate amidotransferase	-56.6
ATP15	YPL271W	ATP synthase epsilon subunit	-12.7
FMT1	YBL013W	Formyl-Methionyl-tRNA Transformylase	-8.5
ESC8	YOL017W	Telomeric and mating-type locus silencing	-8.2
QCR10	YHR001W-A	Ubiqunol-cytochrome c oxidoreductase complex	-7.7
POR1	YNL055C	Outer mitochondrial membrane porin	-4.1
ATP7	YKL016C	ATP synthase D subunit	-3.3
ATP2	YJR121W	F(1)F(0)-ATPase complex beta subunit	-2.2
SER3	YER081W	Catalyzes the first step in serine biosynthesis	1.8
ADH3	YMR083W	Alcohol dehydrogenase isoenzyme	1.9
RTG1	YOL067C	Interorganelle communication	4.7
CRC1	YOR100C	Mitochondrial carnitine carrier	4.8
FOL1	YNL256W	Folic acid synthesis	5.7
IDH1	YNL037C	Mitochondrial isocitrate dehydrogenase	5.8
RGR1	YLR071C	Transcription regulation of diverse genes	6.6
IDH2	YOR136W	NAD+-dependent isocitrate dehydrogenase	6.6
OAC1	YKL120W	Mitochondrial oxaloacetate carrier	7.0
SIR3	YLR442C	Silencing at HML, HMR, and telomeres	9.9
GCR1	YPL075W	Positive regulator of the enolase	11.7
RSC58	YLR033W	Chromatin remodeling complex subunit	12.1
RPM2	YML091C	Mitochondrial precursor tRNAs	44.7

Table 2.3 Differentially expressed genes in PSY3642

Under aerobic conditions, the model predicted a slight growth disadvantage in PSY3642 attributed to diversion of carbon flux away from biomass into formic acid synthesis (Figure 2.1A) whereas under anaerobic conditions the predicted growth rates for the two strains were equivalent (data not shown). Consistent with these predictions, under aerobic conditions we observed a substantial growth defect (0.17 h^{-1} versus 0.26 h^{-1}) (Figure 2.2E). Under anaerobic conditions, the two strains had comparable growth rates (0.23 h^{-1} versus 0.24 h^{-1}) (Figure 2F). One simple explanation for the exacerbated growth defect observed in PSY3642 was toxicity. However, the addition of high concentrations of extracellular formic acid up to 100mM were well tolerated and did not affect relative rates of growth or cell lysis in either strain (data not shown).

Engineering endogenous C1 metabolism induces the retrograde response: To gain insight into potential transcription mechanisms underlying increases in formic acid secretion, we performed gene expression analysis comparing PSY3639 and PSY3642 by competitive hybridization to whole-genome cDNA microarrays (see Methods). Bioinformatics analysis of these data implicated specific transcription responses resulting from manipulating C1 metabolism for formic acid production. An initial assessment of genes with the highest differential expression (some as high as 56-fold) revealed a varied transcription response involving disparate biological processes including glucose repression (*GCR1* and *RGR1*), mitochondrial function (*ATP15*, *FMT1*, *RPM2*, and *QCR10*), telomere maintenance (*ESC8*, *RSC58*, and *SIR3*), and C1 metabolism (*ADE4*, *ATP2*, *ATP7*, *FOL1*, *FMT1*, and *POR1*) (Table 2.3). Significant gene ontology enrichments were identified for enzymes primarily involved in mitochondrial-associated reactions ($p = 0.01$) including TCA metabolic processes ($p = 0.02$) and oxidative phosphorylation ($p = 0.005$). These results are generally consistent with glycine-induced transcription changes observed for C1 metabolic enzymes (Gelling et al., 2004).

A

B

Motif	Factors	Function
	Azf1	Mitochondrial transcription
	Cat8, Sip4	Glucose catabolism
	Gcr1	
	Mig1-3	
	Rgt1	
	Dal80	Nitrogen metabolism
	Dal81, 82	
	Hap1	Respiratory gene expression
	Hap2-5	
	Rox1	
	Ecm22, Upc2	Sterol synthesis
	Mot3	
	Abf1	Chromatin organization
	Rap1	
	Rep1	
	Rtg1, 3	Retrograde signaling

Figure 2.3 Expression analysis of PSY3642.

Figure 2.3, continued. Expression analysis of PSY3642. Hierarchical clustering of PSY3642 and other mutant strain expression profiles are represented as a clustering diagram (dendogram) (A). For clarity, only a portion (approximately one-third) of the complete dendogram is shown. The cluster subtrees are ordered and displayed non-randomly such that similar expression profiles appear closer together. Labels for several functional categories reflect the tendency for strains with comparable genetic deficiencies to exhibit similar patterns of gene expression (Hughes et al., 2000). PSY3642 clusters close to strains with mutations affecting chromatin function, MAP kinase signaling, and mitochondrial function (see text for details). DNA-binding motifs ($p < 10^{-10}$) within the promoters of activated genes in PSY3642 are represented as sequence logos (left column) along with cognate transcription factors (middle column) and identified regulatory roles (right column) (B).

To identify similar patterns of expression in other mutant backgrounds, we compared the gene expression profile of PSY3642 with the compendium generated by Hughes et al. (2000). Using hierarchical clustering of these data we identified patterns of expression that were similar to the profile observed in PSY3642 (Figure 2.3A). Each of these profiles was associated with a particular conditional experiment. We chose to limit our analysis to gene expression profiles generated from gene deletions. PSY3642-similar expression patterns were associated with specific gene mutations affecting chromatin function and general transcription regulation, MAP kinase signal transduction (*swi6*, *sst2*, *dig1*, and *dig2*), mitochondrial function (*rip1*, *qcr2*, *kim4*, etc.), and cell wall (*gas1*, *anp1*, *fks1*, etc.) and ergosterol biosynthesis (*erg2* and *erg3*). Interestingly, single deletions in *SSN6*, *RPD3*, or *TUP1*—components of a well-characterized transcription silencing complex—resulted in the differential expression of over 180 genes (Green and Johnson, 2004; Smith and Johnson, 2000), many of which were also differentially expressed in PSY3642. Motif enrichments in the promoters of up-regulated genes implicated transcription factors associated with several biological processes (Figure 2.3B). Included in this set were Rtg1 and Rtg3, two transcription factors that mediate mitochondria-to-nucleus retrograde signaling in response to severe mitochondrial dysfunction.

Retrograde signaling is typically associated with the petite phenotype caused by loss of mitochondrial DNA (ρ^0) (Butow and Avadhani, 2004). Cells sense mitochondrial dysfunction and implement systemic changes in gene expression to compensate for mitochondria-associated

metabolic deficiencies (Liu and Butow, 2006). In ρ^0 yeast, Rtg1-mediated retrograde signaling is exclusively post-translational: phosphorylated Rtg1 translocates to the nucleus without any change in the abundance of *RTG1* transcript itself; however, in PSY3642 expression of *RTG1* is significantly up-regulated (Table 2.3). This suggests an alternative (transcription) mode of Rtg1-mediated retrograde signaling that is absent in ρ^0 yeast.

Using mitochondria-specific vital stains we confirmed the mitochondrial defect implied by induced expression of *RTG1* and transcriptional induction of retrograde responsive genes. Whereas PSY3639 resembled wild type yeast with regard to mitochondrial abundance, morphology, and membrane potential, analysis of PSY3642 revealed a heterogeneous population of cells that were depleted in functional mitochondria (Figure 2.4). Microscopic analysis revealed a dramatic reduction in both the total number of mitochondria and their associated respiratory capacity (comparing Figures 2.4A and C to Figures 2.4B and D). This effect was primarily evident in PSY3642, with strains PSY3640 and PSY3641 resembling the parent strain PSY3639 (Supplemental Figure 1). These data, along with experiments that show elevated formic acid secretion exclusively under oxygenated growth conditions (Figure 2.2) indicate that mitochondrial function and respiratory capacity are severely diminished but not completely abolished in PSY3642.

ρ^0 yeast up-regulate several genes in order to provide stoichiometric amounts of oxaloacetate and acetyl-CoA to drive the TCA cycle in the presence of respiratory deficiency (Epstein et al., 2001). Several of these genes are also up-regulated in PSY3642 including *PYC1*, *OAC1*, and *CRC1*, as well as the NAD-dependent TCA cycle enzymes *IDH1* and *IDH2* (Table 2.3). As the FBA model predicted very little flux through the TCA cycle, a retrograde responsive increase in flux through these reactions represents an unanticipated (and unforeseeable) adaptation to manipulating C1 metabolism. We constructed an *rtg1* mutation in the PSY3642 genetic background (PSY3650) and observed a 25% increase in formic acid secretion (Table 2.4).

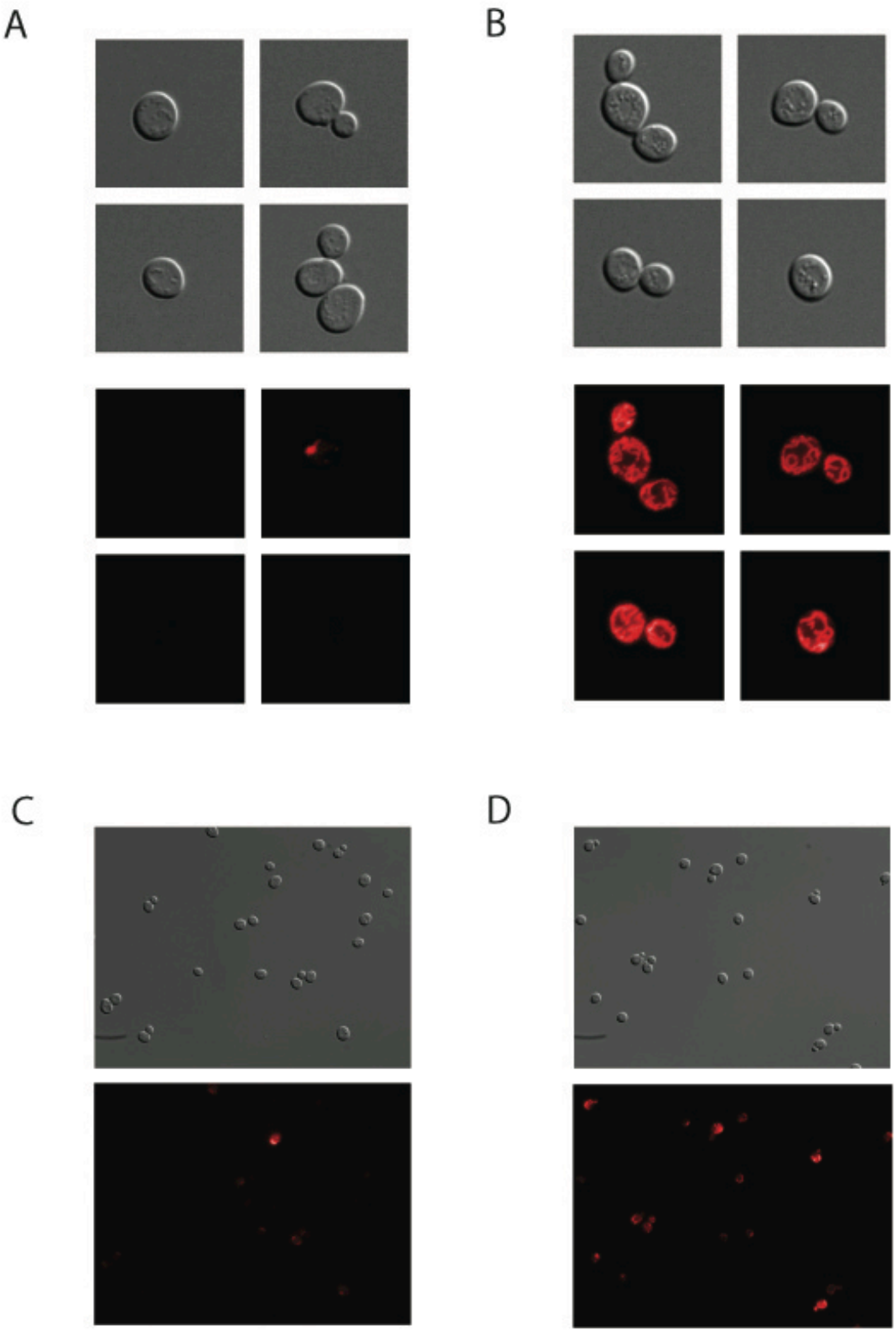


Figure 2.4 Phenotypic analysis of PSY3642 reveals mitochondrial dysfunction.

Figure 2.4, continued. Phenotypic analysis of PSY3642 reveals mitochondrial dysfunction. Total mitochondria were labeled in PSY3642 (A) and PSY3639 (B) and single-cells were imaged at 100X magnification with DIC (top panels) and epifluorescence (bottom panels). Oxidation of Mitotracker® CM-H2XRos was measured to indicate respiratory capacity in PSY3642 (C) and PSY3639 (D). Cells were imaged at 40X magnification with DIC (top panels) and epifluorescence (bottom panels). Exposure times were 40ms in all cases.

In the presence of severe respiratory deficiency increased TCA cycle flux would be coupled to increased flux through NAD-dependent reactions (for example through increased expression of *IDH1* and *IDH2*). As a result the cell would require some biochemical mechanism to regenerate NAD and maintain redox homeostasis. ρ^0 yeast up-regulate the expression of glycerol-3-phosphate and alcohol dehydrogenase enzymes (Gpd2 and Adh1-7) as part of the retrograde response. These pathways reoxidize NADH in the absence of competent oxidative capacity (Epstein et al., 2001). In PSY3642 there is no significant change in expression of *GPD2*, *ADH1*, or *ADH4-7*. *ADH3*—a mitochondrial ethanol-acetaldehyde redox shuttle (Bakker et al., 2000) is induced 2-fold; however, we observed no detectable difference in ethanol production between PSY3642 and PSY3639 (Table 2.4). To test the hypothesis that flux through NAD-dependent reactions increase in PSY3642 we measured intracellular NADH/NAD ratios directly (see Methods). We observed a significant increase in intracellular NADH relative to NAD, which was eliminated in *rtg1* mutants (Figure 2.5). Together, these results support the hypothesis of an Rtg1-mediated increase in flux to the TCA cycle, which increases the NADH/NAD ratio and diverts organic substrate away from C1 metabolism and formic acid biosynthesis.

Strain	Formic Acid (mM)[a]	Ethanol (mM)[a]
PSY3639	0.01 $\pm$ 0.005	155 $\pm$ 19
PSY3642	0.16 $\pm$ 0.004	178 $\pm$ 14
PSY3650	0.2 $\pm$ 0.01	NA

Table 2.4 Quantification of Rtg1-mediated formic acid secretion

[a] Data are given as the average of three biological replicates $\pm$ SD.

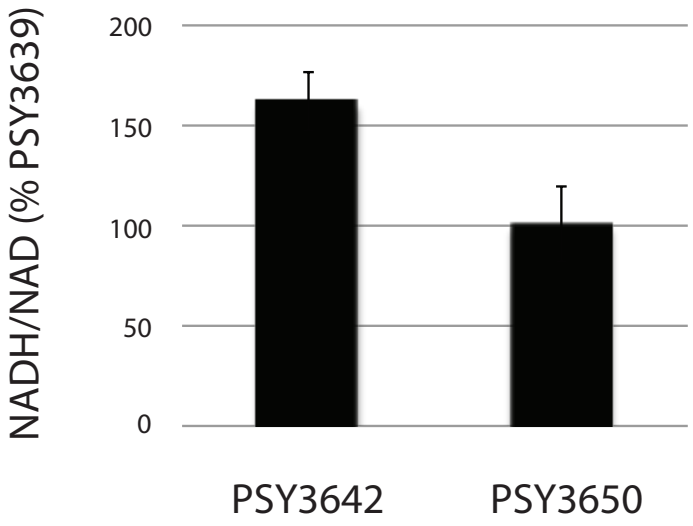


Figure 2.5 Increase in NADH/NAD in PSY3642. NADH/NAD ratios for PSY3642 and PSY3650 are given as percent of PSY3639. Data represent the average of four biological replicates $\pm$ SD. Compared to PSY3639, NADH accumulated significantly in PSY3642 ($p = 0.001$) but not in PSY3650 ($p = 0.46$). Statistical significance was assessed by testing the null hypothesis $m = 100\%$ versus the alternative $m > 100\%$ assuming normally distributed sample data.

Increased flux to formic acid is modulated by coordinated expression of C1 pathway enzymes: According to our pathway analysis of C1 metabolism, the enzymes associated with Eq. (1) are both necessary and sufficient to catalyze the full oxidation of glucose into formic acid given proper regulatory constraints that serve to channel flux through this pathway. Transcription co-regulation of the upstream portion of Eq. (1) (glycolysis) is well characterized. By combining pathway analysis and transcription data, we tested the hypothesis that downstream reactions occurring after the glycolytic branchpoint PGDH are also subject to co-regulation. We sought to identify endogenous transcription programs responsible for increased C1 metabolic flux in PSY3642.

A closer look at relative mRNA abundance for enzymes involved in formic acid biosynthesis revealed an interesting pattern of endogenous differential expression (Figure 2.6B). Relative to the

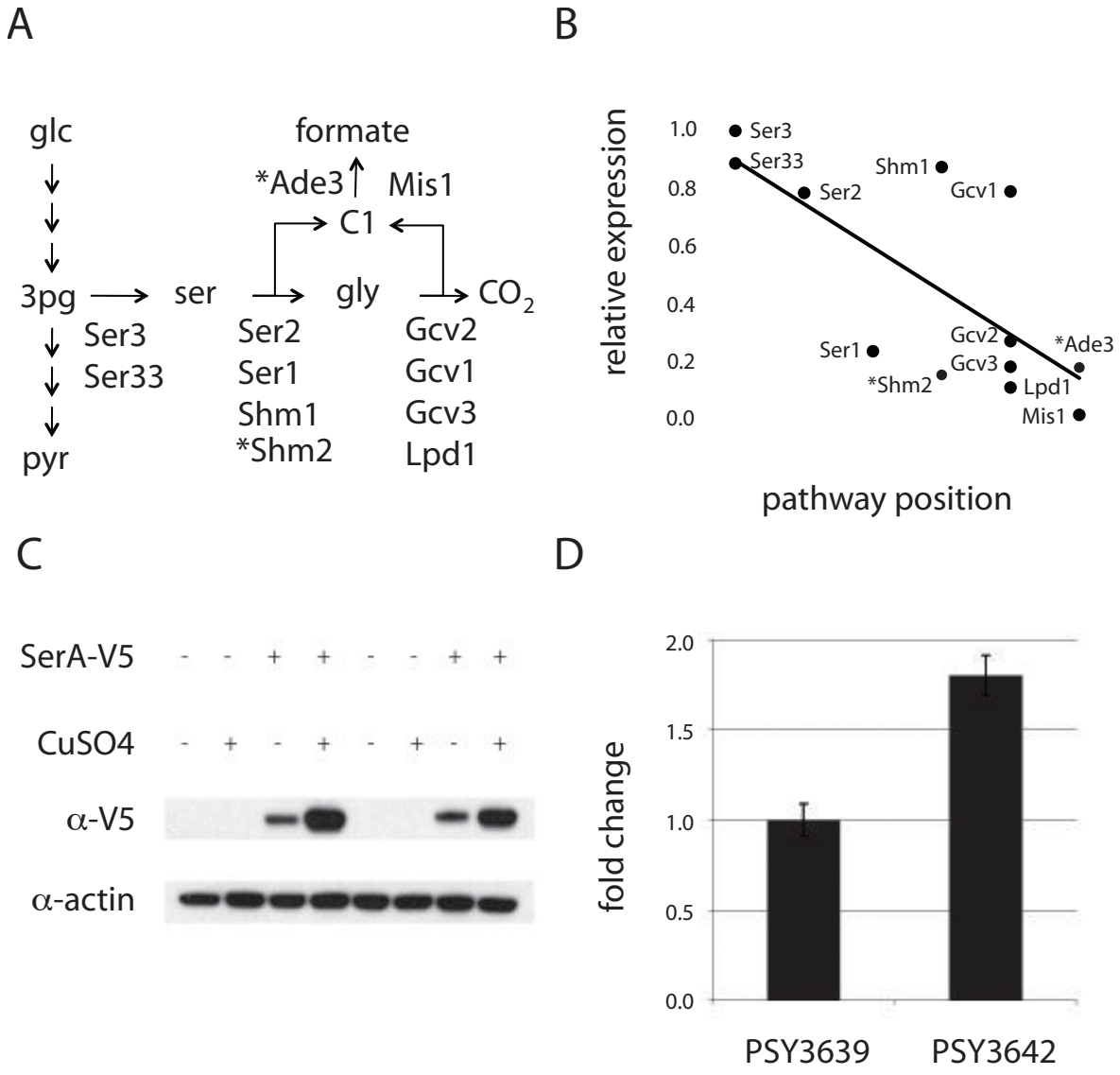


Figure 2.6 Flux to formic acid is controlled by PGDH and the coordinated expression of all pathway enzymes. The subnetwork of serine-glycine-formate biosynthesis with reaction arrows that represent composites of several biochemical interconversions labeled with relevant enzymes (A). The cytoplasmic enzymes involved in converting serine into formic acid (Shm2 and Ade3) are denoted with asterisks. Differential expression of enzymes in (A) were normalized to Ser3 and plotted according to their pathway position (B). Multisite modulation for enzymes in the formic acid biosynthetic pathway is indicated by linear regression (black line, $R^2 = 0.52$) and corrected Spearman ranked order correlation (rho = -0.72, p = 0.02). Levels of the SerA fusion protein (relative to actin) were unaffected by genetic background and induction conditions (C). Expressing SerA in PSY3642 caused a significant increase in formic acid secretion under inducing conditions (D). Each transformed strain was normalized to the empty vector control plasmid pRS410a. Data in (D) represent the average of three biological replicates $\pm$ SD. We assessed statistical significance using a paired two-tailed t-test (p = 0.002).

branchpoint isozyme Ser3 (PGDH) differential expression of downstream enzymes is correlated to their relative position in the pathway, with the lowest expression change associated with the terminal enzyme Mis1. Linear regression indicates that 52% of the variance in differential expression of C1-associated enzymes is explained simply by their relative topological position in the reaction pathway ($p = 0.02$). Interestingly, this pattern of differential expression implicates a recognized mechanism of endogenous metabolic control termed multisite modulation, where coordinated expression of several enzymes modulates flux through entire metabolic pathways (Fell, 1997). This provides experimental evidence that C1 enzymes constitute a module of pathway function in PSY3642.

The transcription data we obtained suggested that increased expression of Ser3 was causally associated with increased flux to formic acid (Table 2.3). Generally, overexpressing branchpoint enzymes rarely affects their associated flux to any significant degree. This is due, in part, to feedback inhibition by downstream metabolites (Fell, 1997). Indeed PGDH is allosterically inhibited by serine and, as a consequence, it has very low flux control in mammalian tissue (Thompson et al., 2005; Bell et al., 2004; Fell and Snell, 1988). To test the hypothesis that PGDH may control formic acid synthesis in PSY3642, we generated a plasmid for inducible overexpression of the catalytic domain of *E. coli* SerA (pRS410a), a well-characterized functional homologue of yeast Ser3. Expression levels of the fusion protein were not affected by strain-specific genetic backgrounds or transcriptional induction (Figure 2.6C). Consistent with previous experiments in mammalian tissues, SerA had no effect on flux to formic acid in PSY3639; however, SerA expression in PSY3642 resulted in an 86% increase in extracellular formic acid concentration (Figure 2.6D). This result is consistent with endogenous induction of Ser3 in PSY3642 and suggests that PGDH has a positive flux control coefficient in this strain.

DISCUSSION

Our goal in this work was to test an FBA-based strategy for engineering C1 metabolism in yeast to increase endogenous formic acid production and describe cellular mechanisms responsible for regulating this important metabolic pathway. FBA and gene knockout simulations identified a non-intuitive combination of genes (*ALT2*, *FUM1*, and *ZWF1*), which individually had no obvious role in formic acid biosynthesis (Figure 2.1). Based on the model predictions, we constructed the quintuple knockout *alt2 fdh1 fdh2 fum1 zwf1* and showed significant oxygen-dependent formic acid secretion in the engineered strain, both during log-phase (Figure 2.2A) and stationary-phase growth (Figures 2.2C and D). Further, maximum formic acid secretion required all five enzyme deletions predicted by the model (Figure 2.2B). Although formic acid is an essential intracellular metabolite, it is not secreted in detectable levels in wild type yeast (Blank et al., 2005; McNeil et al., 1996). Thus, our results demonstrate the successful application of an FBA-based strategy for microbial production of formic acid under aerobic growth conditions. More generally, these data support the predictive potential for this approach in deriving strategies aimed at engineering non-fermentative metabolic pathways that integrate across subcellular compartments in eukaryotic microbes.

To gain insights into the regulatory events that result from manipulating formic acid biosynthesis, we supplemented model predictions and validation experiments with gene expression and phenotypic analyses. From these data we identified (i) a significant response involving multiple cellular processes (Table 2.3 and Figure 2.3); (ii) activation of retrograde signaling, mitochondrial dysfunction, and diminished respiratory capacity (Table 2.4 and Figures 2.4 and 2.5); and (iii) regulatory events that lead to the coordinated expression of enzymes involved in C1 metabolism (Figure 2.6). Upon close inspection, these data indicate specific mechanisms of metabolic regulation that result from unanticipated adaptive responses to manipulating C1 metabolic flux.

Several lines of evidence strongly suggest activation of the retrograde response in PSY3642 cells. In terms of global gene expression pattern, PSY3642 is most similar to strains with mutations in genes that are directly involved in retrograde signaling (Figure 2.3A). These include Rpd3, Ssn6, and Tup1, a well-characterized co-repressor complex (Malave and Dent, 2006); and Yat2, a carnitine acetyl-CoA transferase involved in transporting activated acetate into respiratory deficient mitochondria (Liu and Butow, 2006; Epstein et al., 2001; Swiegers et al., 2001). Interestingly, Yat2 was identified in our original *in silico* knockout screen for enzymes that affect C1 metabolic flux (Table 2.1).

In petite cells Ssn6-Tup1 is converted from a co-repressor complex into a co-activator, which up-regulates gene expression through direct interaction with Rtg3 (Conlan et al., 1999), one of three transcription factors primarily responsible for retrograde responsive gene activation in yeast (Rothermel et al., 1997). DNA binding motifs for Rtg1 and Rtg3 are overrepresented in the promoters of activated genes in PSY3642 (Figure 2.3D) while *RTG1* expression itself is increased almost 5-fold (Table 2.3). These results strongly implicate retrograde regulatory transcription factors as specific modulators of gene activation and metabolic activity in PSY3642. Consistent with this hypothesis are data showing Rtg1-dependent NADH accumulation and limitations in formic acid biosynthesis (Figure 2.5 and Table 2.4).

While *zwf1* mutants have no mitochondrial defects (Blank et al., 2005), Wu and Tzagoloff (1987) speculated that the petite-like phenotype of *fum1* mutants is caused by decreased concentrations of intramitochondrial amino acids, which limits the production of respiratory chain components. However, loss of Fum1 alone does not account for retrograde signaling in PSY3642, as changes in *fum1* single mutants are limited; only about 20 genes are affected, none of which encode typical retrograde responsive TCA cycle enzymes (McCammon et al., 2002). Furthermore, although Gelling et al. (2004) showed that C1 metabolic flux changes significantly affect the expression status

of respiratory chain components, mitochondrial deficiency and retrograde signaling were not reported as a consequence. From these results we conclude that enzyme deletions predicted by our model cause systemic metabolic changes that modulate C1 flux and activate retrograde signaling in PSY3642. Specifically, *alt2*-associated loss of cytoplasmic alanine transaminase activity—either alone or in combination with *fum1* or *zwf1* mutations—may significantly diminish aminogenic capacity leading to full activation of the retrograde response. Alternatively, retrograde signaling may result from reduced intracellular concentrations of specific retrograde inhibitors such as glutamate, glutamine, or ammonia (Butow and Avadhani, 2004; Dilova et al., 2004; Tate and Cooper, 2003; Crespo et al., 2002).

Multisite modulation is an important mechanism of metabolic control where changes in pathway flux result from the coordinated expression of multiple pathway enzymes in relative proportion to their distance from the main pathway branchpoint (Thomas and Fell, 1996; Fell, 1997). Accumulating evidence obtained in yeast (Niederberger et al., 1992), mammals (Werle et al., 2005; Vogt et al., 2002; Hillgartner et al., 1995; Waterman and Simpson, 1989; Brownie and Pedersen, 1986), and plants (Anterola et al., 1999; Quick et al., 1991; Stitt et al., 1991) suggests that multisite modulation may be a general design principle employed by cells to regulate metabolic flux in vivo (Wildermuth, 2000). Indeed, using formal pathway analysis, we identified a module of pathway function corresponding to downstream reactions in C1 metabolism (Eq. (1)) and observed a pattern of differential expression in PSY3642 that strongly suggested multisite modulation of enzymes in this pathway (Figure 2.6B).

A predominant challenge in constraint-based analysis of metabolic networks involves discovering and incorporating relevant cellular regulatory processes (Stelling et al., 2004; Price et al., 2003). Because regulatory events are conditional and often characterized by their dynamic nature,

they are difficult to predict under the assumptions of conventional FBA (Jamshidi and Palsson, 2008). Without the benefit of complete knowledge of gene regulation it is useful to combine systems-level modeling with experimental data generated post hoc for the purposes of design and discovery in complex metabolic systems. By combining formal pathway analysis, FBA, and experimentation, we were able to identify and exploit specific modes of endogenous regulation to increase C1 metabolic flux and engineer a formic acid producing strain of yeast.

MATERIALS AND METHODS

Constraint-based modeling and *in silico* gene deletion: The validated, genome-scale metabolic model *Saccharomyces cerevisiae i*ND750 previously described by Duarte et al., 2004 was used to model the fully compartmentalized yeast metabolic network. The FBA optimization problem was formulated as described previously (Varma and Palsson, 1994) in the GNU MathProg language and solved with custom generated C code (available upon request) implementing the GNU Linear Programming (LP) Kit (GLPK) available at ftp://aeneas.mit.edu/pub/gnu/glpk.

Specifically, we defined quasi steady-state conditions using the yeast stoichiometric matrix S and unknown flux vector v

$$S \cdot v = 0,$$

with maximization of growth rate (μ) as the objective function for FBA:

$$\max \mu$$
$$s.t. v_{\min,i} \le v_i \le v_{\max,i},$$

Thermodynamic constraints in *i*ND750 are derived from the KEGG database (http://www.genome.jp/kegg/) and associated with individual reactions to define v_i^{ub} and v_i^{lb}, which are the upper and lower bounds of each reaction i. We modeled 20 mmol gDW^{-1} h^{-1} constant glucose uptake; nutrient uptake fluxes appropriately constrained to simulate synthetic complete media including the addition of serine (Figure 2.1A, Supplementary Table S2.1); oxygen uptake flux

that was either fixed to 0 mmol gDW^{-1} h^{-1} or left unconstrained to simulate anaerobic or aerobic growth conditions, respectively; and internal fluxes constrained to $\{0, \infty\}$ or left unconstrained to simulate irreversible or reversible reactions, respectively. It is important to note that multiple optimal solutions are possible in which the objective function constraint is satisfied (Phalakornkule et al., 2001; Lee et al., 2000). We sampled alternative optimal solutions for the mutant strains predicted to increase flux to formic acid by relaxing the directionality constraint of individual reactions (Lee et al., 1997). Results from this analysis indicated some flexibility in the formic acid biosynthetic pathway; however, the three mutations *alt2*, *fum1*, and *zwf1* were consistently associated with significant increases in formic acid secretion, with a minimum secretion rate of 53.65 mmol gDW^{-1} h^{-1} (other data not shown).

To identify the maximum theoretical yield for formic acid production we substituted formic acid secretion for biomass in the objective function. All the other constraints were appropriate for external exchange and aerobic growth. Maximizing this objective resulted in Eq. (1), which can be considered the 'Type I' through pathway for formic acid biosynthesis (Shilling et al., 2000).

Several efficient algorithms have been developed for identifying a single solution of optimal gene knockouts (for example, Burgard et al., 2003). We chose an iterative simulation strategy as we were interested in all combinations of gene knockouts predicted to impact C1 metabolism and formic acid production, including solutions that might be considered sub-optimal under formal definitions. By constraining the fluxes of individual sets of ≤ 3 non-essential genes to zero and reevaluating the system using LP we simulated metabolic phenotypes for greater than 4 million gene combinations in reasonable time frames (less than four hours on an x86-64 processor running Linux version 2.6.15-51). Our knockout simulation protocol is summarized with the following pseudocode:

$$p \leftarrow \text{powerset}(\{\text{non-essential genes}\})$$
$$\text{foreach combo} \in p : |p| \leq 3$$
$$\quad \text{foreach c} \in \text{combo}$$
$$\quad\quad v_c^{ub} = 0.0$$
$$\quad\quad v_c^{lb} = 0.0$$

Yeast strains and plasmids: PSY3642 was derived from the *fdh1 fdh2* parental strain PSY3639 (Overkamp et al., 2002) by iterative gene replacement (Guldener et al., 2002). Briefly, LoxP-KanMX gene deletion cassettes for *ALT2*, *FUM1*, and *ZWF1* were generated by PCR using primers with 45bp of flanking homology and pUG6 as template (Guldener et al., 1996). $KanMX^+$ transformants were selected on YPD plates containing 200mg/l G418 (geneticin). After confirming integration by single colony PCR, G418 sensitivity was reestablished for subsequent gene replacement by expressing Gal4-Cre from pSH65 (Guldener et al., 2002) and selecting transformants on YPD plates containing 50mg/l phleomycin. Correct excision of LoxP-KanMX was confirmed by single colony PCR.

The biobrick assembly method (Knight, 2003; Phillips and Silver, 2006) was used to generate the expression plasmid pRS410a consisting of (ordered 5' to 3'): the yeast *CUP1* promoter; the yeast Kozak sequence; the catalytic domain of SerA (cloned by PCR from *E. coli* genomic DNA); in-frame fusion of the V5 epitope; stop codon; and the yeast *ADH1* terminator. The flanking biobrick restriction sites XbaI and SpeI were used to subclone the expression fragment into pRS410 (Addgene). Yeast transformants were selected on YPD plates containing 200mg/l G418 and cultured with the same concentration of drug in synthetic complete media containing 2% glucose and 0.3 mM $CuSO_4$. Primer sequences used in this study are available as Supplementary Table S2.3.

Metabolite determination: Extracellular formic acid and ethanol measurements were made using spectrometric enzymatic assays at 340nm according to the manufacturer's specifications (R-Biopharm). Yeast cultures were grown in synthetic complete media with 2% glucose under batch

conditions in baffled (aerobic) or round bottom (anaerobic) Erlenmeyer shaker flasks. Anaerobic cultures were grown in sparged media under nitrogen gas.

Samples for intracellular metabolite measurements were prepared using several methods as conceptual frameworks (Canelas et al., 2008; Sporty et al., 2008; Visser et al., 2004; and Lange et al., 2001). Briefly, 20ml samples were quickly drawn at a log phase (cell density of 0.4-0.6 OD_{600}) and immediately quenched in 32ml cold 60% (v/v) methanol. A frozen binary solution of 60% (v/v) ethanol was used to maintain the yeast samples and quenching solution at -40ºC. Two subsequent washes were performed using the cold quenching solution, followed by cell lysis at 4ºC using a glass bead beater in the presence of 75% (v/v) ethanol to precipitate proteins. Samples were lyophilized overnight, resuspended in 100ul anaerobic water, and centrifuged twice before use. The supernatant was stored at -80ºC until processing via HPLC.

HPLC measurements were performed using a Waters 2695 HPLC separation module fitted with a Luna C18 5µ column, 250x4.5mm (Phenomenex). Samples of 75-85µl were injected and eluted at a flow rate of 1ml/min, starting with 100% mobile phase buffer A and gradually increasing to 100% mobile phase buffer B (Di Pierro et al., 1995). Specifically, the relative fraction of buffer B in the mobile phase was increased at a rate of 15%/min until 60%, at 0.6%/min until 80% during which the majority of separation occurred, and at 20%/min until 100% was reached. Buffer A contained 10mM tetrabutylammonium hydroxide, 10mM KH_2PO_4, and 0.25% methanol at pH 7.0 (Di Pierro et al., 1995). Buffer B contained 2.8mM tetrabutylammonium hydroxide, 100mM KH_2PO_4 and 30% methanol at pH 5.5 (Di Pierro et al., 1995). NAD was analyzed at 260nm, and NADH was analyzed at 340nm using a photodiode array detector (Waters 996) (Di Pierro et al., 1995 and Sporty et al., 2008). Standard curves of specific metabolites were performed to enable quantification.

Gene expression profiling and analysis: PolyA mRNA was obtained in biological triplicate by trizol extraction (Invitrogen) from early log-phase (OD_{600} = 0.4) yeast grown aerobically.

cDNA was generated from PSY3642 and PSY3639 by reverse transcription; differentially labeled with Cy3 or Cy5, respectively (one sample was processed with the labeling reversed to minimize artifacts introduced by incorporation bias); and hybridized to whole-genome cDNA microarrays (http://www.microarray.ca/). Array scans were analyzed with GenPix software and Rosetta Resolver (complete data are available as Supplementary Table S2.4).

Gene ontology enrichment was obtained using GoStat (http://gostat.wehi.edu.au/) with Benjamini false discovery to correct for multiple hypothesis testing. The query set for enzyme annotation analysis was limited to those gene identities with represented reactions in *i*ND750 for which flux values were calculated.

Hierarchical clustering was applied to genome-wide expression profiles of PSY3642 (compared to PSY3639) and other genetic knockout strains (excluding overexpression and drug treatment conditions) in the compendium described by Hughes et al., 2000. Routines were implemented in R (http://www.r-project.org/) with a Euclidean distance metric and Ward's minimum variance clustering algorithm (Murtagh, 1985).

We used YEASTRACT-DISCOVERER (http://www.yeastract.com/) to find transcription factor binding motif enrichments in the promoter regions of genes with significant activation ($p < 0.01$) greater than 2-fold (Monteiro et al., 2008 and Teixeira et al., 2006).

Phenotypic analysis: Mitotracker® Red CMXRos and Mitotracker® CM-H_2XRos (Invitrogen) were used to stain mitochondria. CMXRos selectively stains mitochondria, and fluoresces in the red portion of the spectrum. CM-H_2XRos is a reduced version of CMXRos, and only fluoresces when oxidized in respiring mitochondria (Ludovico et al., 2002). Overnight yeast cultures were resuspended in YEPD media containing either 1μM of CMXRos or 3μM CM-H_2XRos, and incubated at 30ºC for 30 minutes prior to imaging.

Antibodies and Western blotting: PVDF membranes were blocked with phosphate buffered saline containing 0.25% Tween-20 (PBST) and 5% nonfat dry milk; probed with mouse monoclonal primary antibodies against actin (Chemicon) or V5 (Sigma) and appropriate HRP-conjugated secondary antibody (Jackson); washed with PBST; and developed with enhanced chemiluminescence substrate (Amersham).

SUPPLEMENTAL FIGURES AND TABLES

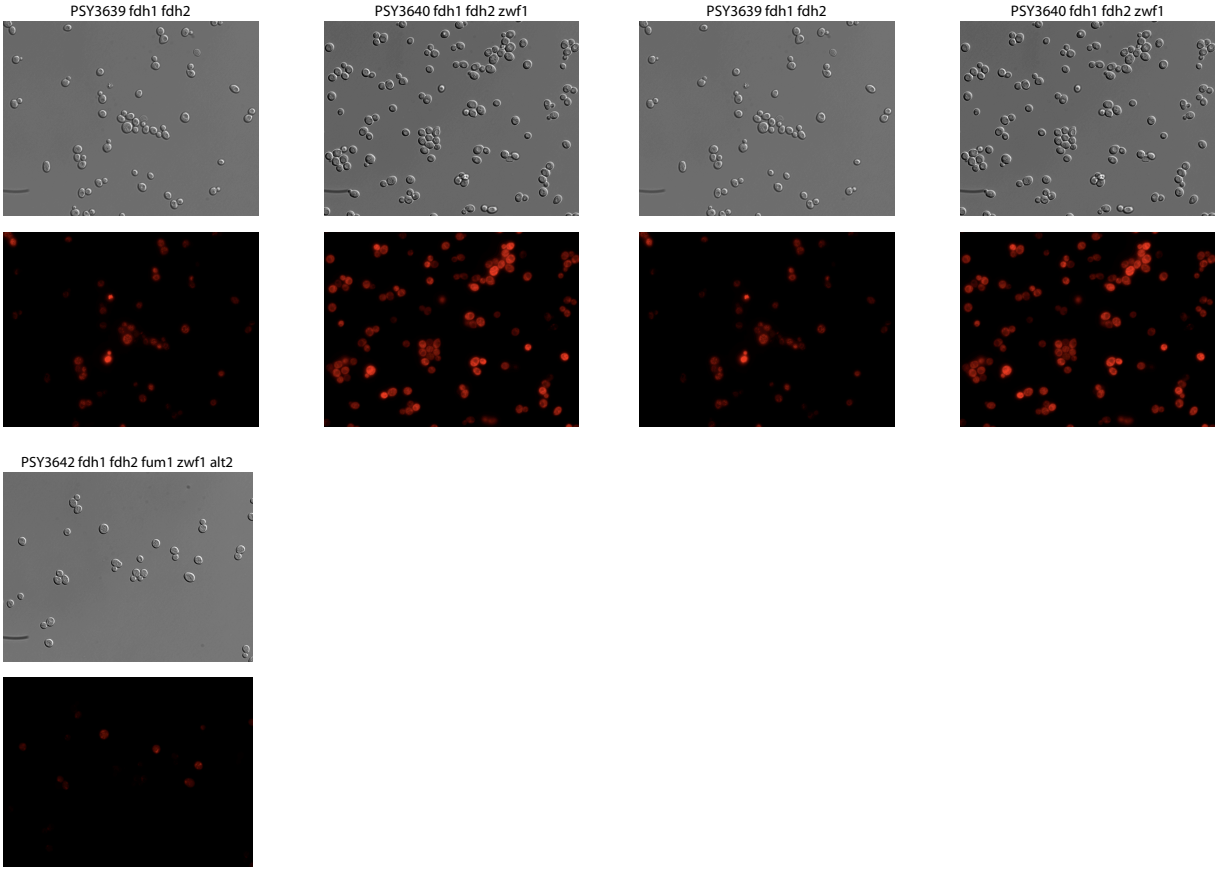


Figure 2.7 Comparison of Mitotracker staining of strains described in this study. For each strain, the upper image is DIC at 60x magnification, and the lower image is the Mitotracker signal from the same field. Materials and methods are identical to the Mitotracker CMXRos protocol described in the main text.

Tables 2.S1 to 2.S4 are available online at
http:/www.genetics.org/ cgi/content/full/genetics.109.105254/DC1.

ACKNOWLEDGEMENTS

We thank Christina Agapakis and Jake Wintermute for comments on the manuscript. This work was supported in part by an NIH training grant to C. J. K., an NIH Cell and Developmental Biology Training Grant (GM07226) to P. M. B., and in part by a grant from the Harvard University Center for the Environment (HUCE).

REFERENCES

Alper H, Moxley J, Nevoigt E, Fink GR, and Stephanopoulos G. (2006) Engineering yeast transcriptional machinery for improved ethanol tolerance and production. *Science*, **314**, 1565-1568.

Anterola AM, van Rensburg H, van Heerden PS, Davin LB, and Lewis NG. (1999) Multi-site modulation of flux during monolignol formation in loblolly pine (*Pinus taeda*). *Biochem Biophys Res Commun*, **261**, 652-657.

Bakker BM, Bro C, Kotter P, Luttik MAH, van Dijken JP, and Pronk JT. (2000) The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J Bacteriol*, **182**, 4730-4737.

Becker SA, and Palsson BO. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*, **4**, e1000082.

Bell JK, Grant GA, and Banaszak LJ. (2004) Multiconformational states in phophoglycerate dehydrogenase. *Biochemistry*, **43**: 3450-3458.

Birkmann A, Zinoni F, Sawers G, and Brök A. (1987) Factors affecting transcriptional regulation of the formate-hydrogen-lyase pathway of Escherichia coli. *Arch Microbiol*, **148**, 44-51.

Blank LM, Kuepfer L, and Sauer U. (2005) Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol*, **6**, R49.

Bro C, Regenberg B, Forster J, and Nielsen J. (2006) In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng*, **8**, 102-111.

Brownie AC and Pedersen RC. (1986) Control of aldosterone secretion by pituitary hormones. *J Hypertens Suppl*, **4**, S72-S75.

Burgard AP, Pharka P, and Maranas CD. (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*, **84**, 647-657.

Butow, RA and Avadhani, NG. (2004) Mitochondrial signaling: the retrograde response. *Mol Cell*, **14**, 1-15.

Canelas, A. B., W. M. van Gulik, and J. J. Heijnen. 2008. Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions. *Biotechnol Bioeng,* **100**, 734-43.

Christensen KE and MacKenzie RE. (2006) Mitochondrial one-carbon metaboism is adapted to the specific needs of yeast, plants and mammals. *BioEssays*, **28**, 595-605.

Conlan SR, Gounalaki N, Hatis P, and Tzamarias D. (1999) The Tup1-Cyc8 protein complex can shift from a transcriptional co-repressor to a transcriptional co-activator. *J Biol Chem*, **1**, 205-210.

Cordier H, Mendes F, Vasconcelos I, and Francois JM. (2007) A metabolic and genomic study of engineered *Saccharomyces cerevisiae* strains for high glycerol production. *Metab Eng*, **9**: 364-378.

Crespo JL, Powers T, Fowler B, and Hall MN. (2002) The TOR-controlled transcription activators *GLN3*, *RTG1*, and *RTG3*, are regulated in response to intracellular levels of glutamine. *Proc Natl Acad Sci USA*, **99**, 6784-6789.

Di Pierro, D., B. Tavazzi, C. F. Perno, M. Bartolini, E. Balestra, R. Calio, B. Giardina, and G. Lazzarino. 1995. An ion-pairing high-performance liquid chromatographic method for the direct simultaneous determination of nucleotides, deoxynucleotides, nicotinic coenzymes, oxypurines, nucleosides, and bases in perchloric acid cell extracts. *Anal Biochem*, **231**, 407-12.

Dilova I, Aronova S, Chen JC-Y, and Powers T. (2004) Tor signaling and nutrient-based signals converge on Mks1p phosphorylation to regulate expression of Rtg1p-Rtg3p-dependent target genes. *J Biol Chem*, **45**, 46527-46535.

Duarte NC, Herrgård MJ, and Palsson BO. (2004) Reconstruction and validation of Saccharomyces cerevisiae *i*ND750, a fully compartmentalized genome-scale metabolic model. *Genome Res*, **14**, 1298-1309.

Edwards JS and Palsson BO. (2000) The *Eschericia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Nat Acad Sci USA*, **97**, 5528-5533.

Epstein CB, Waddle JA, Hale W 4th, Dave V, Thornton J, Macatee TL, Garner HR, and Butow RA. (2001) Genome-wide responses to mitochondrial dysfunction. *Mol Biol Cell*, **12**, 297-308.

Fell DA. (1997) Frontiers in metabolism: understanding the control of metabolism. London: *Portland Press*.

Fell DA and Snell K. (1988) Control analysis of mammalian serine biosynthesis. Feedback inhibition on the final step. *Biochem J*, **256**: 97-101.

Friend SH. (2000) Functional discovery via a compendium of expression profiles. *Cell*, **102**, 109-126.

Forster J, Famili I, Palsson B, and Nielsen J. (2003) Large-scale evaluation of *in silico* gene knockouts in *Saccharomyces cerevisiae*. *OMICS*, **7**, 193-202.

Fu TF, Rife JP, Schirch V. (2001) The role of serine hydroxymethyltransferase isozymes in one-carbon metabolism in MCF-7 cells as determined by ^{13}C NMR. *Arch Biochem Biophys*, **393**, 42-50.

Geertman JM, van Maris AJ, van Dijken JP, and Pronk JT. (2006) Physiological and genetic engineering of cytosolic redox metabolism in *Saccharomyces cerevisiae* for improved glycerol production. *Metab Eng*, **8**, 532-542.

Gelling CL, Piper MD, Hong SP, Kornfeld GD, Dawes IW. (2004) Identification of a novel one-carbon metabolism regulon in *Saccharomyces cerevisiae*. *J Biol Chem*, **279**, 7072-7081.

Green SR and Johnson AD. (2004) Promoter-dependent roles for the Srb10 cyclin-dependent kinase and the Hda1 deacetylase in Tup1-mediated repression in *Saccharomyces cerevisiae*. *Mol Biol Cell*, **15**, 4191-4202.

Guldener U, Heck S, Fiedler T, Beinhauer J, and Hegemann JH. (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res*, **24**, 2519-2524.

Hefetz A and Blum MS. (1978) Biosynthesis of formic acid by the poison glands of formicine ants. *Science*, **201**, 454-455.

Herrero O, Ramon D, and Orejas M. (2008) Engineering the *Saccharomyces cerevisiae* isoprenoid pathway for de novo production of aromatic monoterpenes in wine. *Metab Eng*, **10**, 78-86.

Hillgartner FB, Salati LM, and Goodridge AG. (1995) Physiological and molecular mechanisms involved in nutritional regulation of fatty acid synthesis. *Physiol Rev*, **75**, 47-76.

Hjersted, J. L., Henson, M. A. and Mahadevan, R. (2007) Genome-scale analysis of *Saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture. *Biotechnol Bioeng*, **97**, 1190-1204.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennet HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stephaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, and

Ishida N, Saitoh S, Ohnishi T, Tokuhiro K, Nagamori E, Kitamoto K and Takahashi H. (2006) Metabolic engineering of *Saccharomyces cerevisiae* for efficient production of pure L: -(+)-lactic acid. *Appl Biochem Biotechnol*, **131**, 795-807.

Jamshidi N and Palsson BO. (2008) Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*, **4**, 171.

Kastanos EK, Woldman YY, and Appling DR. (1997) Role of mitochondrial and cytoplasmic serine hydroxymethyltransferase isozymes in *de novo* purine synthesis in *Sacharomyces cerevisiae*. *Biochemistry*, **36**, 14956-14964

Kizer L, Pitera DJ, Pfleger BF, and Keasling JD. (2008) Application of functional genomics to pathway optimization for increased isoprenoid production. *Appl Environ Microbiol*, **74**, 3229-3241.

Klamt S and Stelling J. (2002) Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep*, **29**, 233-236.

Knight T. (2003) Idempotent vector design for standard assembly of biobricks. http://hdl.handle.net/1721.1/21168. In DSpace. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group.

Lange, H. C., M. Eman, G. van Zuijlen, D. Visser, J. C. van Dam, J. Frank, M. J. de Mattos, and J. J. Heijnen. 2001. Improved rapid sampling for in vivo kinetics of intracellular metabolites in Saccharomyces cerevisiae. *Biotechnol Bioeng*, **75**, 406-15.

Lee J, Goel A, Ataii MM, Domach MM. (1997) Supply-side analysis of growth of *Bacillus subtilis* on glucose-citrate medium: feasible network alternatives and yield optimality. *Appl Environ Microbiol*, **63**, 710-718.

Lee S, Phalakornkule C, Domach MM, Grossmann IE. (2000) Recursive MILP model for finding all alternate optima in LP models for metabolic networks. *Comp Chem Eng*, **24**, 711-716.

Leonhartsberger S, Korsa I, and Bock A. (2002) The molecular biology of formate metabolism in enterobacteria. *J Mol Microbiol Biotechnol*, **4**, 269-276.

Liu Z and Butow RA. (2006) Mitochondrial retrograde signaling. *Annu Rev Genet*, **40**, 159-185.

Ludovico P, Rodrigues F, Almeida A, Silva MT, Barrientos A, and Corte-Real M. (2002) Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **13**, 2598-2606.

Malave TM, and Dent SY. (2006) Transcriptional repression by Tup1-Ssn6. *Biochem Cell Biol*, **84**, 437-443.

McCammon MT, Epstein, CB, Przybyla-Zawislak B, McAlister-Henn L, and Butow RA. (2003) Global transcription analysis of Krebs tricarboxylic acid cycle mutants reveals an alternating pattern of gene expression and effects on hypoxic and oxidative genes. *Mol Biol Cell*, **14**, 958-972.

McNeil JB, Bognar AL, and Pearlman RE (1996) In vivo analysis of folate coenzymes and their compartmentalization in Saccharomyces cerevisiae. *Genetics*, **142**, 371-381.

Milliken CE and May HD. (2007) Sustained generation of electricity by the spore-forming, Gram-positive, *Desulfitobacterium hafniense* strain DCB2. *Appl Microbiol Biotechnol*, **73**, 1180-1189.

Monteiro PT, Mendes N, Teixeira MC, d'Orey S, Tenreiro S, Mira N, Helios P, Francisco AP, Carvalho AM, Lourenco A, Sa-Correia I, Oliveria AL, and Freitas AT. (2008) YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucl Acids Res*, **36**, D132-D136.

Motter, A. E., Glubahce, N., Almass, E. and Barabasi, A. L. (2008) Predicting synthetic resuces in metabolic networks. *Mol Syst Biol*, **4**, 168.

Murtagh F. (1985) Multidimensional clustering algorithms. Wuerzburg: *Physica-Verlag*.

Niederberger P, Prasad R, Miozzari G, and Kacser H. (1992) A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast. *Biochem J*, **287**, 473-479.

Oh Y-K, Palsson BO, Park SM, Schilling CH, and Mahadevan R. (2007) Genome-scale reconstruction of metabolic network in *Bacillus subtillis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*, 10.1074/jbc.M703759200.

Overkamp KM, Kötter P, van der Hoek R, Schoondermark-Stolk S, Luttik MA, van Dijken JP, and Pronk JT. (2002) Functional analysis of structural genes for NAD(+)-dependent formate dehydrogenase in Saccharomyces cerevisiae. *Yeast*, **19**, 509-520.

Papin J, Price N, Edwards J, and Palsson B. (2002) The genome-scale metabolic extreme pathway structure in *Haemophilus influenze* shows significant network redundancy. *J Theor Biol*, **215**, 67-82.

Peters-Wendisch P, Stolz M, Etterich H, Kennerknecht N, Sahm H, and Eggeling L. (2005) Metabolic engineering of *Corynebacterium glutamicum* for L-serine production. *Appl Environ Microbiol*, **71**, 7139-7144.

Phalakornkule C, Lee S, Zhu T, Koepsel R, Ataai MM, Grossmann IE, Domach MM. (2001) A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering. *Metab Eng*, **3**, 124-137.

Phillips I and Silver PA. 2006, A new biobrick assembly strategy designed for facile protein engineering. http://hdl.handle.net/1721.1/32535. In DSpace. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group.

Piper MD, Hong SP, Ball GE, and Dawes IW. (2000) Regulation of the balance of one-carbon metabolism in *Saccharomyces cerevisiae*. *J Biol Chem*, **275**, 30987-30995.

Price N., Reed JL, Papin JA, Wiback SJ, and Palsson BO. (2003) Network-based analysis of metabolic regulation in the human red blood cell. *J Theor Biol*, **225**, 185-194.

Quick WP, Schurr U, Sheibe R, Schulze E-D, Rodermel SR, Bogorad L, and Stitt M. (1991) Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in tobacco transformed with 'antisense' *rbcS*. I. Impact on photosynthesis in ambient growth conditions. *Planta*, **183**, 542-555.

Resendis-Antonio O, Reed RJ, Encarnacion S, Collado-Vides J, and Palsson BO. (2007) Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comp Biol*, **3**, e192.

Rothermel BA, Thornton JL, and Butow RA. (1997) Rtg3p, a basic helix-loop-helix/leucine zipper protein that functions in mitochondrial-induced changes in gene expression, contains independent activation domains. *J Biol Chem*, **32**, 19801-19807.

Schuster S, Fell DA, Dandekar T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, **18**: 326-332.

Shuster S, Klamt S, Weckwerth W, Moldenhaurer F, and Pfeiffer T. (2002) Use of network analysis of metabolic systems in bioengineering. *Bioproc Biosyst Eng*, **24**, 363-372.

Shiba Y, Paradise EM, Kirby J, Ro DK, and Keasling JD. (2007) Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of isoprenoids. *Metab Eng*, **9**, 160-168.

Schilling CH, Letsher D, Palsson BØ. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, **203**, 229-248.

Smith RL and Johnson AD. (2000) Turning genes off by Ssn6-Tup1: a conserved system of transcriptional repression in eukaryotes. *Trends Biochem Sci*, **25**, 325-330.

Sporty, J. L., M. M. Kabir, K. W. Turteltaub, T. Ognibene, S. J. Lin, and G. Bench. 2008. Single sample extraction protocol for the quantification of NAD and NADH redox states in Saccharomyces cerevisiae. *J Sep Sci*, **31**, 3202-11.

Stein I, Peleg Y, Even-Ram S, and Pines O. (1994) The single translation product of the FUM1 gene (fumarase) is processed in the mitochondria before being distributed between the cytosol and the mitochondria in Saccharomyces cerevisiae. *Mol Cell Biol*, **14**, 4770-4778.

Stelling J and Gilles ED. (2004) Mathematical modeling of complex regulatory networks. *IEEE Trans Nanobioscience*, **3**, 172-179.

Stitt M, Quick WP, Schurr U, Schulze E-D, Rodermel SR, and Bogorad L. (1991) Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in tobacco transformed with 'antisense' *rbcS*. II Flux-control coefficients for photosynthesis in varying light, CO_2, and air humidity. *Planta*, **183**, 555-566.

Subramanian M, Qiao WB, Khanam N, Wilkins O, Der SD, Lalich JD, Bognar AL. (2005) Transcriptional regulation of the one-carbon metabolism regulon in *Saccharomyces cerevisiae* by Bas1p. *Mol Microbiol*, **57**, 53-69.

Swiegers JH, Dippenaar N, Pretorius IS, and Bauer FF. (2001) Carnitine-dependent metabolic activities in *Saccharomyces cerevisiae*: three carnitine acetyltransferases are essential in a carnitine-dependent strain. *Yeast*, **18**, 585-595.

Tate JJ and Cooper TG. (2003) Tor1/2 regulation of retrograde gene expression in *Saccharomyces cerevisiae* derives indirectly as a consequence of alterations in ammonia metabolism. *J Biol Chem*, **38**, 36924-36933.

Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, and Sa-Correia I. (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nuc Acids Res*, **34**, D446-D451.

Thomas S and Fell DA. (1996) Design of metabolic control for large flux changes. *J Theor Biol*, **182**, 285-298.

Thompson JR, Bell JK, Bratt J, Grant GA, and Banaszak LJ. (2005) Vmax regulation through domain and subunit changes. The active form of phosphoglycerate dehydrogenase. *Biochemistry*, **44**: 5763-5773.

Varma A and Palsson BO. (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology*, **12**, 994-998.

van Maris AJ, Winkler AA, Porro D, van Dijken JP, and Pronk JT. (2004) Homofermentative lactate production cannot sustain anaerobic growth of engineered *Saccharomyces cerevisiae*: possible consequence of energy-dependent lactate export. *Appl Environ Microbiol*, **70**, 2898-2905.

Visser, D., G. A. van Zuylen, J. C. van Dam, M. R. Eman, A. Proll, C. Ras, L. Wu, W. M. van Gulik, and J. J. Heijnen. 2004. Analysis of in vivo kinetics of glycolysis in aerobic Saccharomyces cerevisiae by application of glucose and ethanol pulses. *Biotechnol Bioeng*, **88**, 157-167.

Vogt AM, Nef H, Schaper J, Poolman M, Fell DA, Kübler W, and Elsässer A. (2002) Metabolic control analysis of anaerobic glycolysis in human hibernating myocardium replaces traditional concepts of flux control. *FEBS Lett*, **24**, 245.

Waks Z and Silver PA. (2009) Engineering a synthetic dual organism for hydrogen production. *Appl Environ Eng*, **75,** 1867-1875.

Waterman MR and Simpson ER. (1989) Regulation of steroid hydroxylase gene expression is multifactorial in nature. *Recent Prog Horm Res*, **45**, 155-163.

Werle M, Kreuzer J, Höfele J, Elsässer A, Ackermann C, Katus HA, and Vogt AM. (2005) Metabolic control analysis of the Warburg-effect in proliferating vascular smooth muscle cells. *J Biomed Sci*, **12**, 827-834.

Wildermuth M. (2000) Metabolic control analysis: biological applications and insights. *Genome Biol*, **1**, 1031.

Wu M and Tzagoloff A. (1987) Mitochondrial and cytoplasmic fumarases in *Saccharomyces cerevisiae* are encoded by a single nuclear gene *FUM1*. *J Biol Chem*, **262**, 12275-12282.

Yoshida, A., Nishimura, T., Kawaguchi, H., Inui, M., and Yukawa, H. (2005) Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing *Escherichia coli* strains. *Appl Environ Microbiol*, **71**, 6762-6728.

Zelle, RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JM, van Dijken JP, Pronk JT, and van Maris AJ. (2008) Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Appl Environ Microbiol*, **74**: 2766-2777.

Chapter 3

Strategies for biological hydrogen production[1]

ABSTRACT

Hydrogen could potentially serve as a clean-burning carbon neutral fuel, provided that efficient carbon neutral processes can be developed for hydrogen production. Biological systems can potentially fulfill this need, given that hydrogenase enzymes are able to efficiently catalyze the production of hydrogen without the use of exotic metallic cofactors. In this chapter we present the optimization of synthetic hydrogen production circuits in *Escherichia coli*, and demonstrate their utility in selecting for hydrogenase mutants with improved properties.

PRODUCING HYDROGEN BIOLOGICALLY

A primary factor limiting the usefulness of hydrogen as a carbon neutral fuel is that 95% of current hydrogen production is achieved through the steam reformation of natural gas (Kontur et al.

[1] Portions of the work presented in this chapter were published in the following papers:

1. Agapakis, C. M., Ducat, D.C., Boyle, P.M., Wintermute, E.H., Way, J.C., & Silver, P.A. Insulation of a synthetic hydrogen metabolism circuit in bacteria. *J Biol Eng* **4**, 3 (2010).
    a. Contributions: CMA, DCD, PMB, EHW, JCW, and PAS designed experiments and analyzed data; CMA, DCD, PMB, EHW performed experiments; CMA, DCD, JCW, and PAS wrote the paper.
2. Barstow, B., Agapakis, C. M.*, Boyle, P.M.*, Gerald, G.*, Silver, P.A., & Wintermute, E.H. A synthetic system links FeFe-hydrogenases to essential E. coli sulfur metabolism. *J Biol Eng* **5**, 7–7 (2011). *Equal contribution.
    a. Contributions: EHW designed experiments and drafted the manuscript. EHW, BB, CMA, PMB and GG performed experiments. EHW, BB, CMA, PMB, GG and PAS analyzed data.

Excerpts from these papers will be indicated via footnotes in this chapter. I am also including some experimental data from work that I did that is unpublished, specifically work on a PFOR-based selection system.

2011). Biological hydrogen ("bio-hydrogen") production utilizes reducing power from sugars or photosynthesis rather than fossil fuels. Furthermore, hydrogenase enzymes do not require rare metals to function, in contrast to current industrial scale electrolysis technologies (Aguirre de Carcer et al. 2006). Two major hurdles limit the utility of current biological systems for hydrogen production: low hydrogen yields from common feedstocks, and the oxygen sensitivity of [FeFe]-hydrogenases (Savage et al. 2008). Re-engineering of metabolic pathways and of the hydrogenase itself is required to address these issues.

We have tested several strategies to improve bio-hydrogen production: the elimination of competing metabolic pathways, direct linkage of the hydrogenase to redox carriers (Agapakis et al. 2010), and genetic selections for the directed evolution of oxygen-tolerant hydrogenases(Barstow et al. 2011).

PROPERTIES OF [FeFe]-HYDROGENASES

Hydrogenase enzymes are remarkable catalysts; they facilitate the interconversion of two protons and two electrons to molecular hydrogen, and do so more efficiently than industrial processes which require rare metals such as platinum (Shima et al. 2008). The vast majority of known hydrogenase enzymes are classified into two categories, nickel-iron [NiFe] and iron-iron [FeFe]. As the nomenclature suggests, these two hydrogenase classes differ in the metal content of their active sites (Vignais & Billoud 2007). These hydrogenase families appear to have evolved independently, as they show no similarity in amino acid sequence. The highly related geometries of hydrogenase active sites, and the use of similar ligands such as carbon monoxide and cyanide, suggest that the physical requirements of hydrogen production have resulted in convergent evolution (Shima et al. 2008; Vignais & Billoud 2007).

[FeFe]-hydrogenases can be expressed heterologously in *Escherichia coli* via coexpression with three maturation factors: HydE, HydF, and HydG. HydE and HydF can be expressed as a functional fusion protein, as demonstrated by the HydEF gene in *Chlamydomonas reinhardtii*, effectively reducing the number of required [FeFe]-hydrogenase maturation factors to two: HydEF and HydG (King et al. 2006). In contrast, the maturation of [NiFe] hydrogenases requires at least 12 maturation factors, making heterologous expression of [NiFe] hydrogenases more difficult (Casalot & Rousset 2001; Savage et al. 2008). This limits the portability of synthetic devices that utilize [NiFe] hydrogenases, as the full complement of required maturation factors is unknown.

An additional advantage of [FeFe]-hydrogenases is that they tend to favor hydrogen production under typical physiological conditions, while [NiFe] hydrogenases favor the reverse hydrogen-consuming reaction (Ghirardi et al. 2007). This is a consequence of relationship between the redox potential of the hydrogenase active site and the redox potential of the carrier molecule that shuttles electrons to the hydrogenase. [NiFe] hydrogenases are coupled to NADH, while [FeFe]-hydrogenases are coupled to ferredoxin (Fd) proteins. NADH and Fd have reducing potentials of -320 mV and -420 mV, respectively, with the latter potential being closer to that of an H_2/H+ pair (Ghirardi et al. 2009). Furthermore, [FeFe]-hydrogenases exhibit among the highest activities of known hydrogenases, with catalytic rates exceeding the 1000 nmol H2 mg-1 min-1 estimated to be required for efficient in vivo photosynthetic hydrogen production (Benemann et al. 1973).

Electrons and protons must reach the catalytic H-cluster of [FeFe]-hydrogenases to produce hydrogen. Protons appear to diffuse in solution to protonate residues near the active site, while electrons are shuttled from Fd (Mulder et al. 2011). Fd are characterized by iron-sulfur (FeS) clusters, which is where electrons are stored following reduction reactions. FeS clusters in the hydrogenase itself serve as "wires," allowing electrons to jump from reduced Fd and through FeS clusters to the H-cluster via quantum tunneling (Page et al. 1999). The modular nature of Fd

proteins and FeS clusters has been exploited by synthetic biologists to generate Fd-hydrogenase fusion proteins that more efficiently evolve hydrogen[2] (Agapakis et al. 2010).

It has been demonstrated that Fd wires can be used to directly couple hydrogenases to photosystem I *in vitro* (Ihara et al. 2006). Achieving sustainable hydrogen production by fusing hydrogenases to photosystem I *in vivo* would prevent captured electrons from entering cellular metabolism, maximizing hydrogen production efficiency. Unfortunately, all known [FeFe]-hydrogenases are sensitive to oxygen, which is a byproduct of photosynthetic activity (Armstrong 2009). While [NiFe] hydrogenases are more oxygen tolerant than [FeFe]-hydrogenases, the superior modularity and faster kinetic rates of [FeFe]-hydrogenases make them intriguing candidates for photosystem fusions.

In this chapter, we present two complementary strategies that work towards sustainable biohydrogen production: the optimization of heterologous hydrogenase pathways in *E. coli*, as well as the design of genetic selection systems for oxygen-tolerant [FeFe]-hydrogenases. To study and manipulate the function of hydrogenases *in vivo*, both strategies involve the coupling of hydrogenases to oxidoreductase enzymes via Fd (Figure 3.1). Pyruvate:ferredoxin oxidoreductase (PFOR) reduces Fd, and reduced Fd in turn provides electrons for hydrogen production by the hydrogenase. Sulfite reductase (SIR) oxidizes Fd in the process of producing hydrogen sulfide, reduced Fd for this process can be provided by a hydrogenase running in the reverse direction, consuming molecular hydrogen.

[2] This aspect of the paper is not presented in detail here, as I did not directly contribute to the Fd-hydrogenase fusion experiments.

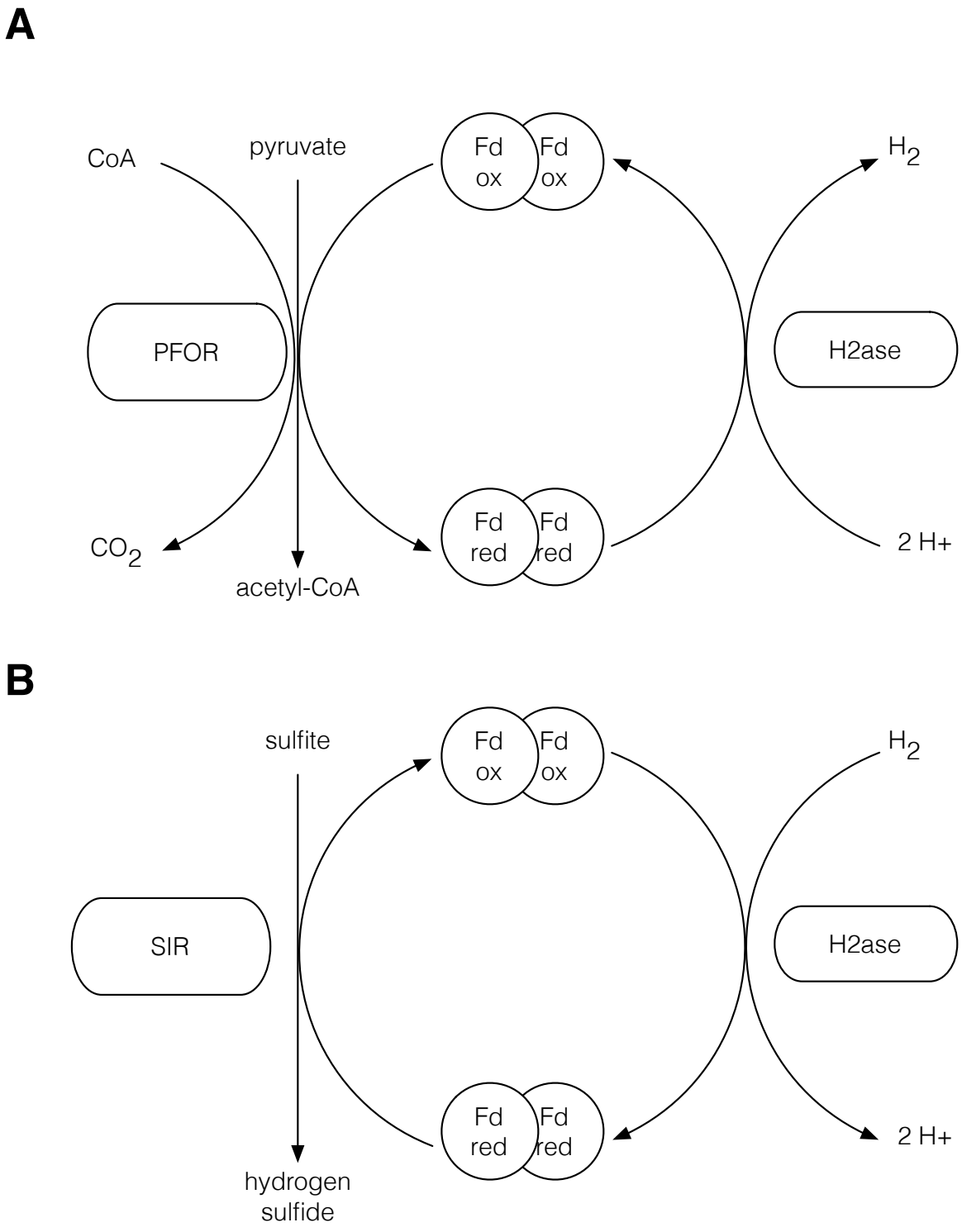


Figure 3.1 Synthetic hydrogenase circuits.

Figure 3.1, continued. Synthetic hydrogenase circuits. (A) Pyruvate:ferredoxin oxidoreductase (PFOR) generates reduced Fd by converting pyruvate to acetyl-CoA. Oxidation of reduced Fd by a hydrogenase produces molecular hydrogen. (B) Sulfite reductase (SIR) reduces sulfite to sulfide via the oxidation of Fd. Consumption of hydrogen by a hydrogenase (H2ase) regenerates reduced Fd.

OPTIMIZING THE PFOR-FERREDOXIN-HYDROGENASE INTERACTION[3]

In vitro *hydrogen production from heterologously expressed hydrogenases*

To create a synthetic electron metabolism circuit with hydrogenase as the terminal electron acceptor, we first investigated the activity of various hydrogenase genes heterologously expressed in the presence of appropriate maturation factors. We adapted a previously established *in vitro* hydrogenase activity assay (King et al. 2006), and measured hydrogen production from crude lysates of bacteria expressing hydrogenases and maturation factors from several species in the presence of a chemical electron donor, methyl viologen. Previous reports have shown that the hydrogenase maturation factors from *C. reinhardtii,* HydEF and HydG, are unstable when heterologously expressed in *E. coli* (King et al. 2006), likely due to the genes' high GC content, while the maturation factors from *Clostridium acetobutylicum* were able to mature [FeFe]-hydrogenases from a wide range of species. Using commercially synthesized, codon optimized maturation factors from *C. reinhardtii* we were able to alleviate the instability of the gene constructs. We found that *in vitro* hydrogen production from the *Clostridium acetobutylicum* hydrogenase was identical when coexpressed with the synthetic maturation factors or with HydE, HydF, and HydG from *C. acetobutylicum* (data not shown). All subsequent experiments were performed using the optimized *C. reinhardtii* maturation factors.

We compared the *in vitro* hydrogen production of [Fe-Fe] hydrogenases from *Clostridium acetobutylicum, Clostridium saccharobutylicum, Chlamydomonas reinhardtii, Shewanella oneidensis,* and *Thermotoga maritima,* all of which are homologous in their catalytic domain (Appendix A, figure S3). All

[3] This section is adapted from Agapakis et al, 2010.

hydrogenases except HydA from *Thermotoga maritima* could be expressed at a high level in *E. coli* (Figure 3.2A), and were functional *in vitro* (Figure 3.2B). Hydrogen levels increased linearly for the first several hours of measurement (data not shown), and we found that levels of hydrogen gas in the headspace after overnight incubation correlated to the relative rate of hydrogenase activity during this linear phase. Our overnight *in vitro* results agree with previous reports of *in vitro* hydrogen production rates, with the hydrogenases from *Clostridium* species producing the highest levels of hydrogen (King et al. 2006). The heterologously expressed hydrogenase from *Shewanella oneidensis* is functional at relatively low levels *in vitro* when both subunits are coexpressed in *E. coli* with maturation factors from *C. reinhardtii*.

The *in vitro* assay is useful to test and compare the activities of heterologously expressed hydrogenase genes, but as the assay uses an exogenous reducing agent, it does not provide information on the electron flux within normal metabolic pathways *in vivo*. To measure electron flux *in vivo* as a function of hydrogen production, hydrogenase activity must be integrated into a functional electron transfer pathway. One well established class of electron donors to hydrogenases are ferredoxins, small soluble proteins that contain iron-sulfur clusters. Construction of a system where hydrogenase activity depends on electron transfer from Fd would allow for comparison to *in vitro* data to provide information on hydrogenase behavior and hydrogenase-ferredoxin interaction dynamics.

In vivo *construction and optimization of a synthetic hydrogen-producing circuit*

To produce hydrogen *in vivo* from glucose, the [FeFe]-hydrogenase was coexpressed with its required maturation factors, Fd, and pyruvate-ferredoxin oxidoreductase (PFOR) from different species. In this heterologous circuit, PFOR oxidizes pyruvate to acetyl-CoA, reducing Fd, which

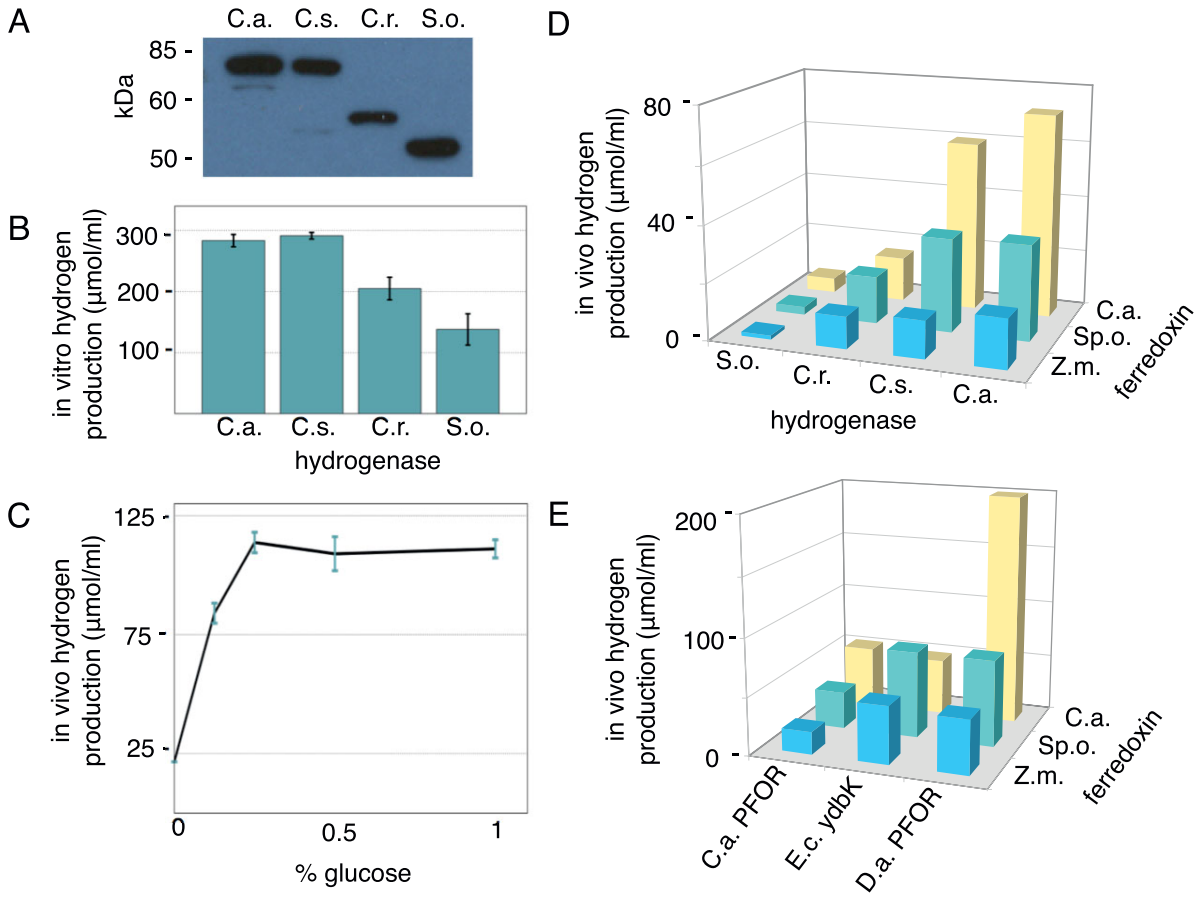


Figure 3.2 Characterization of synthetic hydrogen production pathway. A) Western blot of Strep-II tagged hydrogenase expression. B) *In vitro* hydrogen production from *E. coli* strains expressing various hydrogenases, measured by the methyl viologen *in vitro* assay (King et al. 2006). C.a. = *C. acetobutylicum,* C.s. = *C. saccharobutylicum,* C.r. = *C. reinhardtii*, S.o. = *Shewanella oneidensis.* C) Glucose-dependence of hydrogen production. Here and below, *in vivo* and *in vitro* hydrogen production values are in units of µmol hydrogen/ml of *E. coli* culture, normalized to an OD_{600} of 0.15 unless otherwise stated. Assays were performed in triplicate, with error bars indicating standard deviation. D) *In vivo* hydrogen production from *E. coli* strains expressing all combinations of the four hydrogenases vs. three Fd from *C. acetobutylicum*, *Spinacia olearcea* (Sp.o), and *Zea mays* (Zm). E) *In vivo* hydrogen production from the *C. acetobutylicum* hydrogenase paired with combinations of three Fd and three PFOR genes.

then transfers the electron to the hydrogenase. In normal *E. coli* metabolism, the oxidative breakdown of pyruvate to acetyl-CoA is performed either aerobically by the pyruvate dehydrogenase complex, reducing NAD^+, or anaerobically by pyruvate formate lyase, generating formate (Figure 3.3). PFOR functions in certain anaerobic bacteria and in eukaryotic parasites that possess hydrogenosomes, organelles evolutionarily related to the mitochondrion that generate a proton

gradient through the production of hydrogen gas (Boxma et al. 2005). PFOR is an attractive electron source for a synthetic hydrogen production circuit as overexpression of a putative *E. coli* PFOR homolog, YdbK, increases *in vivo* hydrogen production by heterologously expressed [FeFe]-hydrogenase and Fd (Kalim Akhtar & Jones 2009), PFOR purified from *Clostridium pasteurianum* has been shown to reduce a number of Fd *in vitro* (Moulis & Davasse 1995), and functional PFOR from *Desulfovibrio africanus* has been recombinantly expressed in *E. coli* (Pieulle et al. 1997).

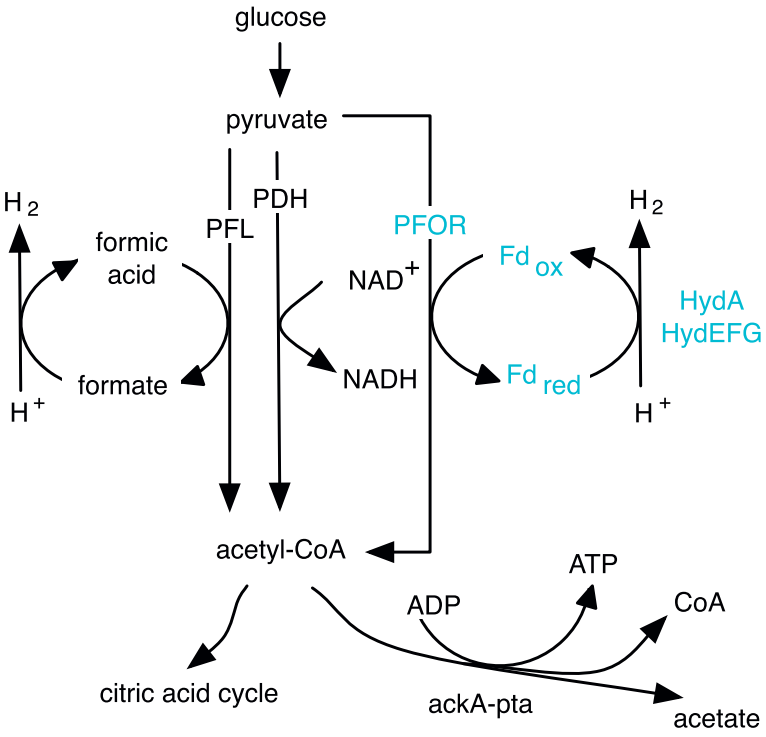


Figure 3.3 Natural and synthetic pyruvate metabolism to acetyl-CoA in *E. coli* through the pyruvate dehydrogenase complex (PDH), pyruvate formate lyase (PFL), and the heterologous PFOR-ferredoxin (Fd)-hydrogenase synthetic pathway. Native enzymes are indicated in black, heterologous enzymes in blue.

Consistent with the establishment of a synthetic electron transport circuit *in vivo*, we observed high levels of glucose-dependent hydrogen production upon coexpression of PFOR, hydrogenase and its maturation factors, and Fd all from *Clostridium acetobutylicum* in an *E. coli* strain lacking endogenous hydrogenases (Δ*hycE,* Δ*hyaB,* Δ*hybC,* Figure 3.2C). Hydrogen production was

again measured after overnight incubation, as we found that hydrogen production *in vivo* from glucose was exhausted after 16 hours (data not shown). We were unable to detect hydrogen production in the parental strain of *E. coli* with the native hydrogenases deleted. Thus, any hydrogen production observed in this strain background must occur via heterologous hydrogenases. Removal of any individual pathway component from the synthetic circuit drastically reduced *in vivo* hydrogen production. However, as has been previously reported, there was a small background level of hydrogen production from expression of hydrogenase and maturation factors alone (Akhtar & Jones 2008). Consistent with previous results (Veit et al. 2008), we found this background hydrogen production was slightly increased upon overexpression of Fd in addition to hydrogenase, indicating that there are *E. coli* proteins capable of reducing both hydrogenases and plant-type ferredoxins, several candidate proteins of which we deleted in the following section (Figure 3.4).

The hydrogenase-ferredoxin-PFOR pathway constitutes a modular system, where each element can be exchanged with homologous genes from different organisms. By coexpressing pathway enzymes from diverse microorganisms, we were able to compare the relative interaction strengths of four hydrogenases, three ferredoxins (Figure 3.2D), and three PFORs (Figure 3.2E). All ferredoxins were able to transfer electrons between PFOR and hydrogenases from different species with varying levels of efficiency.

In vivo hydrogen production from circuits expressing each of the four hydrogenases (*C. acetobutylicum, C. saccharobutylicum, C. reinhardtii,* and *S. oneidensis*) followed the same trend as the *in vitro* experiments, with the highest hydrogen production observed with the clostridial hydrogenases (Figure 3.2D). The relative interaction and electron transfer rates for hydrogenase and Fd were explored by comparing the *in vivo* hydrogen production of circuits made up from all pairwise combinations of the four hydrogenases and Fd from *C. acetobutylicum, Spinacea olearcea,* and *Zea mays* and the PFOR from *C. acetobutylicum* (Figure 3.2D). All hydrogenases produced the highest output

when co-expressed with bacterial type 2-[4Fe-4S] Fd from *Clostridium acetobutylicum,* with a potential of -420mV (Guerrini et al. 2008). Intermediate levels of hydrogen were produced using leaf-type [2Fe-2S]-Fd I from spinach, *S. olearcea* (-420 mV (Yonekura-Sakakibara et al. 2000)) while the homologous root-type Fd III from corn, *Z. mays* (-345 mV (Yonekura-Sakakibara et al. 2000)) led to significantly lower *in vivo* hydrogen levels in all cases. Interestingly, the difference in hydrogen production from circuits expressing bacterial versus plant-type ferredoxins was more significant for hydrogenases from bacterial species. Hydrogenase from *C. reinhardtii*, which naturally pairs with plant-type ferredoxins, produced similar levels of hydrogen when co-expressed with Fd from *C. acetobutylicum* or *S. olearcea* (Figure 3.2D).

The interaction of overexpressed PFOR from *C. acetobutylicum, D. africanus,* or the PFOR homolog YdbK from *E. coli* with the three ferredoxins was compared in a similar fashion in circuits containing the *C. acetobutylicum* hydrogenase (Figure 3.2E). Overexpression of PFOR from *C. acetobutylicum* and YdbK from *E. coli* led to similar levels of hydrogen production, although surprisingly, the highest levels of hydrogen produced from YdbK occurred when it was coexpressed with plant-type Fd from *S. olearcea.* Overall, the highest levels of hydrogen production were seen with the PFOR from *D. africanus,* coexpressed with the hydrogenase and Fd from *C. acetobutylicum.*

Isolation of the hydrogen producing circuit through deletion of competing reactions

Natural biological electron transfer circuits are insulated to prevent electron leaks that can cause damage by creating oxygen radicals and insulated from one another to prevent "short circuiting" (McCord 2000). We sought to insulate our hydrogen producing circuit from competing metabolism to improve levels of hydrogen production and to better understand natural biological pathway isolation, a priority for the design of synthetic metabolic pathways. Although our constructed pathway is made up of genes that are divergent from *E. coli* metabolic enzymes, given the non-specific electrostatic interactions that mediate many Fd interactions (Gou et al. 2006), native iron-

sulfur proteins may interact with the proteins of the heterologous pathway. This is evidenced by the background hydrogen production in strains expressing only heterologous hydrogenases and ferredoxins (Figure 3.4). Deletion of these potentially competing redox interaction partners should improve pathway function. To address these issues, we deleted six genes identified through their homology to plant-type ferredoxins or ferredoxin oxidoreductases that still allowed for viability (*fpr*, flavodoxin:$NADP^+$ reductase (Veit et al. 2008); *ydbK*, the putative PFOR homolog (Veit et al. 2008); *hcr*, an NADH oxidoreductase; *yeaX*, a predicted oxidoreductase; *hcaD*, ferredoxin:NAD^+ reductase; and *frdB*, fumarate reductase. Appendix A, figure S4).

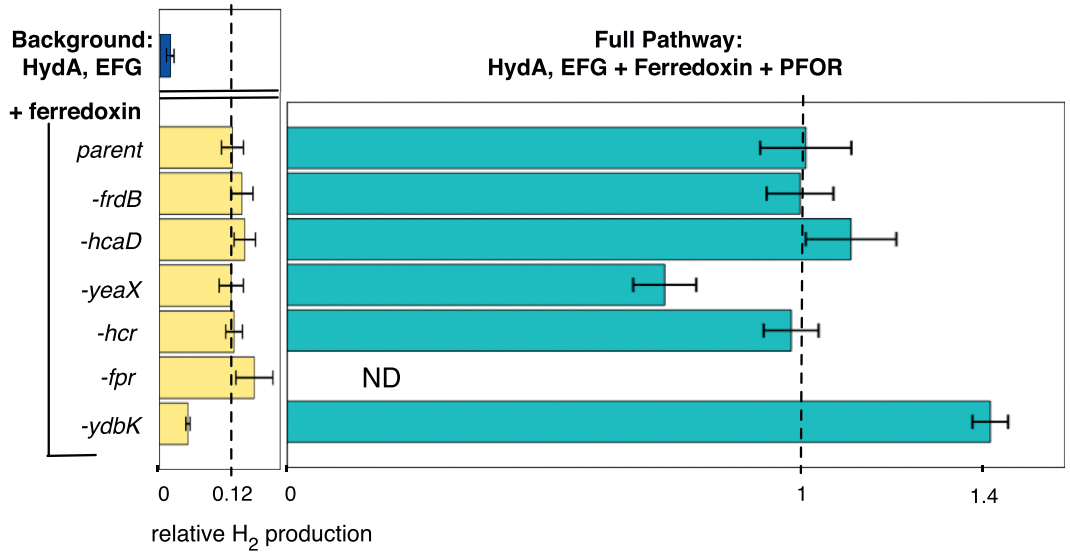


Figure 3.4 Insulation of hydrogenase pathway through deletion of competing reactions. Relative hydrogen production of different knockout strains compared to parent strain (Δ*hycE,* Δ*hyaB,* Δ*hybC*) expressing hydrogenase alone (dark blue bar), hydrogenase and Fd only (yellow bars) or the full PFOR-ferredoxin-hydrogenase pathway (green bars).

These six deletions were tested individually in a Δ*hycE,* Δ*hyaB,* Δ*hybC* background while expressing hydrogenase from *C. acetobutylicum* and maturation factors from *C. reinhardtii,* Fd from *S. olearcea*, with or without co-expression of PFOR from *D. africanus*. Deletion of *fpr* and *ydbK* have been previously shown to slightly decrease the background level of hydrogenase activity *in vivo* (Veit et al. 2008). We found that only the *ydbK* deletion had any significant effect on hydrogen production

compared to the hydrogenase knockouts alone. The background level of hydrogen production from HydA and Fd expressed alone was decreased by half in the *ydbK* deletion strain, whereas hydrogen production from the full pathway with the *D. africanus* PFOR was increased by 1.4 fold (Figure 3.4). This is consistent with our finding that overexpression of *ydbK* led to high levels of electron transfer when co-expressed with Fd from spinach, indicating that endogenous *ydbK* is able to disrupt the synthetic electron transfer pathway.

DESIGN OF A PFOR-BASED GENETIC SELECTION FOR OXYGEN TOLERANT HYDROGENASES[4]

We investigated whether the PFOR-Fd-Hydrogenase circuit (Figure 3.1A) was a suitable pathway for a genetic selection. The rationale for this selection is that cell viability could be coupled to central carbon metabolism, by replacing pyruvate dehydrogenase and pyruvate formate lyase with PFOR. In this manner, PFOR would provide the only route between pyruvate and acetyl-CoA. The reduced Fd generated by PFOR would then be oxidized by an [FeFe] hydrogenase. Thus, the function of the hydrogenase would be required for the maintaining cellular redox balance, and hydrogen production would be an observable output of this process. If cell growth could be coupled to hydrogenase function, this system could be used to select for hydrogenases with improved function; in particular we were interested in identifying oxygen-tolerant hydrogenases.

In parallel, we also developed a genetic selection based on an SIR-Fd-Hydrogenase circuit (Figure 3.1B). For the SIR-based circuit, the hydrogenase is acting in the opposite reaction direction to the PFOR-dependent selection. We hypothesized that if the SIR-dependent and PFOR-dependent selections yielded different types of hydrogenase mutants, then that would indicate a

[4] This work was performed concurrently with the work in Agapakis et al 2010 and Barstow et al 2011. This section describes work that was not included in these publications.

physical mechanistic difference between the forward and reverse hydrogenase reactions. Such a difference could be a consequence of hydrogenase converting aqueous protons to gaseous molecular hydrogen, and vice versa. Unfortunately, our efforts to engineer a PFOR-dependent strain were not successful. The strains that were constructed and tested for the PFOR-based selection are described here. The SIR selection was eventually published as Barstow et al 2011, and is described in the next section.

As with the previously described experiments, our initial strain for the PFOR selection was BL21 *E. coli* with the native hydrogenases knocked out (Δ*hycE,* Δ*hyaB,* Δ*hybC*) (Agapakis et al. 2010). From this strain, we deleted all pyruvate consuming reactions in *E. coli* (Figure 3.5). An *E. coli* with these deletions was previously demonstrated to overproduce pyruvate (Zhu et al. 2008). Zhu et al observed that this strain is only capable of growth on rich media containing acetate, such as TYA medium.

In order to select for oxygen tolerant hydrogenases, all other components of the selective pathway must be oxygen tolerant. The PFOR from *D. africanus* is the only PFOR known to be functional in the presence of oxygen (Pieulle et al. 1997)[5]. As noted previously, we also observed that *da*PFOR produces more hydrogen *in vivo* when coupled with *ca*Fd and *ca*HydA than *ca*PFOR does. Thus, *da*PFOR was used for these experiments. Unless otherwise noted, all strains in this section contain the IPTG-inducible *da*PFOR, *ca*HydA, *ca*Fd, *cr*HydEF,*cr*HydG system described in Agapakis et al 2010.

As with Zhu et al 2008, we knocked out pyruvate related genes *ldhA, pps, poxB,* and *pflB* to increase flux to pyruvate. We also knocked out *fre*, which in early experiments with the SIR-dependent selection had been shown to reduce nonspecific growth (E.H. Wintermute, personal

[5] This is the original reason we cloned *da*PFOR, the improvement in hydrogen yield over the *ca*PFOR was unanticipated.

communication). The pyruvate dehydrogenase genes *aceE* and *aceF* have the largest fitness impact when deleted (Zhu et al. 2008), so we first conducted experiments with the pyruvate dehydrogenase complex intact. Paradoxically, despite no major changes in anaerobic growth on LB (Figure 3.6A), the introduction of PFOR reduced *in vivo* production of hydrogen in this knockout background (Figure 3.6B). The addition of pyruvate and thiamine pyrophosphate (TPP), a substrate of PFOR, had no major effect on improving hydrogen yields (data not shown).

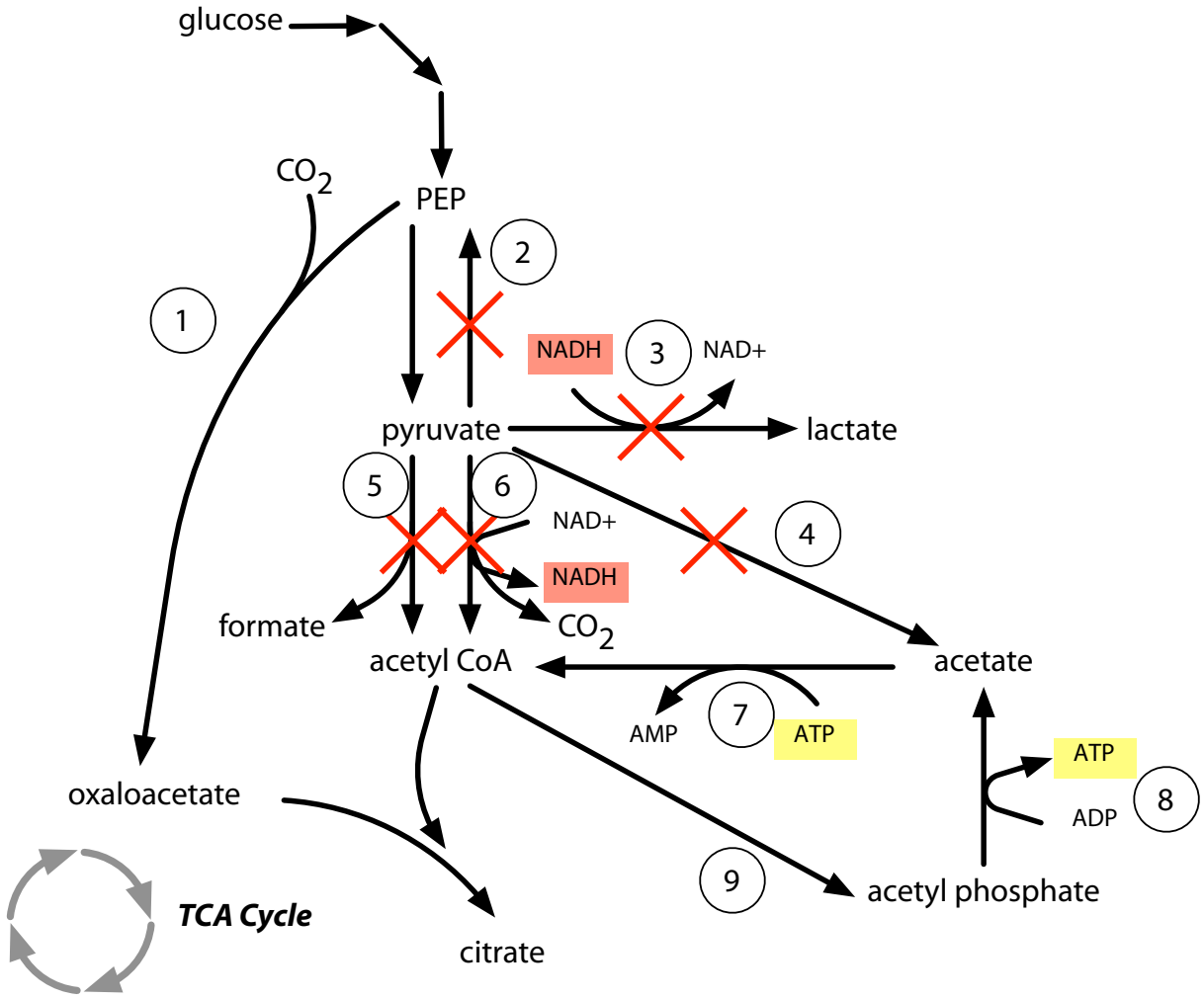


Figure 3.5 Pyruvate metabolism in *E. coli*. Enzymes: 1, PEP carboxylase, 22, PEP synthase, 3, lactate dehydrogenase, 4, pyruvate oxidase, 5, pyruvate formate lyase, 6, pyruvate dehydrogenase complex, 7, acetyl-CoA synthetase, 8, acetate kinase, 9, phosphotransacetylase. Reactions knocked out for the PFOR selection are labeled with a red X. Figure adapted from Zhu et al 2008.

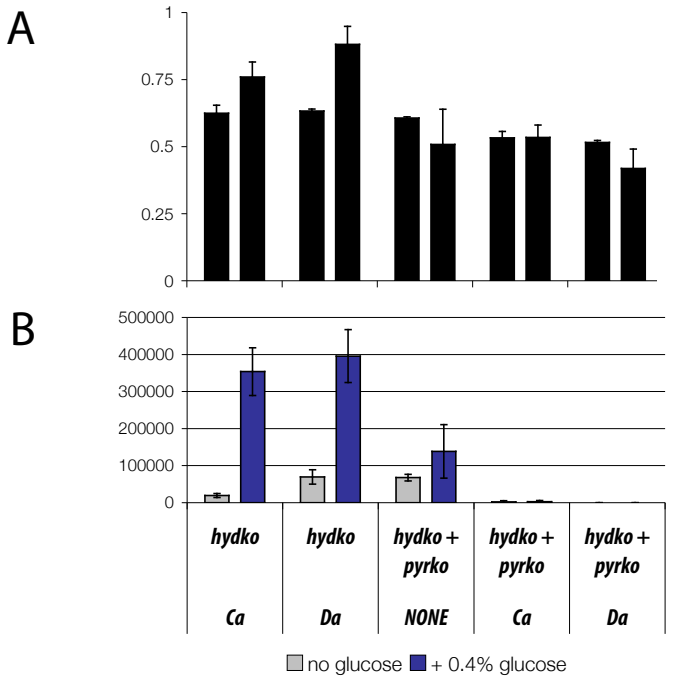


Figure 3.6 Impact of PFOR Expression on *in vivo* hydrogen yields. A) OD 600 of cultures tested for hydrogen production. B) Hydrogen yields from an *in vivo* hydrogen production assay (Agapakis et al. 2010). Ca = *Ca*PFOR, Da = *Da*PFOR, hydko = Δ*hycE* Δ*hybC* Δ*hyaB*, pyrko = hydko + Δ*ldhA* Δ*pps* Δ*poxB* Δ*pflB* Δ*fre*. Y-axis: peak area as detected by gas chromatography (arbitrary units)

Despite these results, we decided to also test pyruvate knockout strains under selection conditions, as in the SIR-dependent selection. Cultures of *ΔldhA Δpps ΔpoxB ΔpflB ΔhycE ΔhybC ΔhyaB Δfre* cells, as well as cultures of cells with only the hydrogenases and one of the two major pyruvate dehydrogenase genes (*aceE* or *aceF*) knocked out. Cells were plated on TYA media, which is rich media containing acetate as a carbon source. In Zhu et all 2008, aerobically culturing pyruvate metabolism knockouts in TYA medium was shown to cause pyruvate accumulation in these strains. *ΔldhA Δpps ΔpoxB ΔpflB ΔhycE ΔhybC ΔhyaB Δfre* were capable of growing both aerobically and anaerobically on TYA media (Table 3.1). Both *ΔaceE* and *ΔaceF* strains, on the other hand, demonstrated poor growth when transformed with PFOR in aerobic conditions. Anaerobically, the growth defect caused by PFOR was sufficient to prevent growth entirely (Table 3.1). Based on these

results, we concluded that our PFOR-based strategy was unsuitable for use as a selection for oxygen-tolerant hydrogenases.

Strain	Growth -PFOR	Growth +PFOR
ΔldhA Δpps ΔpoxB ΔpflB ΔhycE ΔhybC ΔhyaB Δfre	+++ (air)	+++ (air)
	++ (nitrogen)	++ (nitrogen)
ΔhycE ΔhybC ΔhyaB ***ΔaceE***	++ (air)	+ (air)
	++ (nitrogen)	No growth (nitrogen)
ΔhycE ΔhybC ΔhyaB ***ΔaceF***	+++ (air)	+ (air)
	++ (nitrogen)	No growth (nitrogen)

Table 3.1 Growth of pyruvate knockouts on TYA media in different atmospheres.
+ = colonies only, ++ and +++ = lawns.

DESIGN OF AN SIR-BASED GENETIC SELECTION FOR OXYGEN TOLERANT HYDROGENASES[6]

Design for hydrogenase-dependent growth

Figure 3.1B details the synthetic pathway designed for our selection and expressed in *E. coli*. An exogenous FeFe-hydrogenase consumes H_2 and reduces Fd. Fd donates electrons to sulfite reductase for the reduction of sulfite to sulfide. Sulfide serves in the host as an indispensible sulfur source for cysteine biosynthesis.

Fd, an electron-carrying iron-sulfur (Fe-S) protein, is the native redox partner of the best-characterized FeFe-hydrogenases (Demuez et al. 2007). An Fd also receives electrons from photosystem I in plants, suggesting that it could be adapted to mediate light-driven H_2 production. We chose Fd as an intermediate with this future application in mind. A Fd homolog, *fdx*, is found in *E. coli*, where it plays an essential role as a scaffold site for iron-sulfur cluster assembly (Nakamura et

[6] This section is adapted from Barstow et al, 2011.

al. 1999). *E. coli* does not appear to use Fd as an electron carrier in the metabolic network, instead relying on NAD(P)H. We therefore hypothesized that Fd chemistry would be insulated from native metabolism.

The kegg pathway database (Kanehisa et al. 2008) identifies three enzymes that produce essential metabolites using Fd as an electron source. Glutamate synthase, nitrite reductase and SIR activities are all essential for the growth of *E. coli* in minimal medium. The native bacterial enzymes draw electrons from NADPH to produce glutamate, ammonia and sulfide, respectively. The analogous enzymes in plants yield the same products while drawing electrons from Fd, a common redox carrier in those species. While all three products are indispensable for *E. coli* viability, sulfide is consumed in the smallest molar quantity (Feist et al. 2007). We therefore chose to employ SIR, reasoning that the small metabolic requirement would afford more tolerance for suboptimal performance of the heterologous pathway.

SIR is not essential on cysteine-containing rich media, but it becomes essential on selective media containing only oxidized sulfur sources such as sulfate or sulfite. In the absence of the native *E. coli* NADPH-dependent SIR, *cysI*, the synthetic pathway is the only metabolic source of reduced sulfur. If the components of this pathway are insulated from any endogenous electron sources, then hydrogenase activity will also be essential for growth. Increasing O_2 concentrations, by inactivating the hydrogenase, will eventually inhibit the ability of host cells to grow on sulfite. This synthetic pathway therefore enables a genetic selection for hydrogenase mutants with an ability to support growth in high O_2.

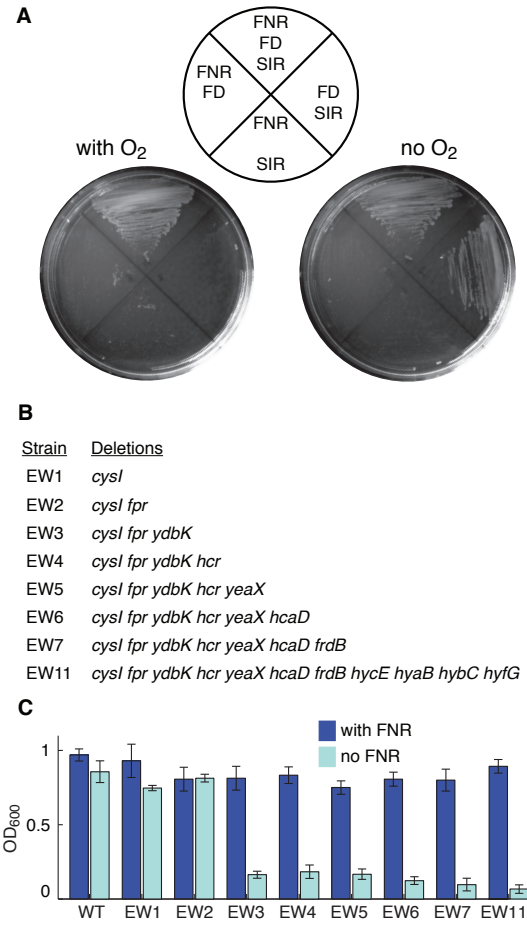


Figure 3.7 Genetic insulation of the synthetic pathway. A) *E. coli* BL21(DE3) *cysI* cells were transformed with plasmids expressing *zm*FNR, *so*FD and *zm*SIR. The *cysI* deletion conveys a requirement for reduced sulfur, which the heterologous pathway supplies. Cells were grown for 24 hours on selective media with our without atmospheric O_2. All three factors were required to rescue growth under aerobic conditions (left plate). Expression of *zm*FNR was not required under anaerobic conditions (right plate), indicating that *so*FD was receiving electrons from another source. B) Genetic deletions were targeted to eliminate potential endogenous redox partners for ferredoxin, therefore linking sulfide production specifically to a synthetic electron source. Also deleted were the catalytic subunits of each native hydrogenase, ensuring that only exogenous hydrogen was present in our system. C) Each deletion strain was transformed with *so*FD, *zm*SIR and either *zm*FNR or an empty plasmid. Growth was measured after 18 hours at 37°C under strict anaerobic conditions, as described in the methods. Sequential deletions reduced the nonspecific anaerobic background growth, with the largest effect produced by the deletion of *ydbK*. The final deletion strain, EW11, showed no growth defect in rich media and was used in all later experiments.

Synthetic ferredoxin-dependent sulfite reduction

We first sought to establish that the native *E. coli* BL21(DE3) SIR could be replaced with a Fd-dependent pathway (Figure 3.7A). Deletion of the *cysI* SIR did not impair growth on rich media

(LB), but eliminated growth on selective media with sulfate as the sole source of metabolizable sulfur. Expression of corn (*Zea mays*) sulfite reductase (*zm*SIR) alone, or together with spinach Fd (*so*FD), failed to rescue growth. This indicated that the *E. coli* host provides no interacting source of reduced Fd or ferredoxin-reductase activity under these conditions.

We then provided a heterologous source of reduced Fd in the form of corn-derived (*Zea mays*) ferredoxin-$NADP^+$ reductase (*zm*FNR). This enzyme links *so*FD to the endogenous NADPH pool by catalyzing redox exchange between the two electron carriers (Yonekura-Sakakibara et al. 2000). FNR requires no maturation factors and is unaffected by O_2; therefore it serves as a hydrogenase-independent control source of electrons for our pathway. Expression of *zm*FNR with *so*FD and *zm*SIR rescued growth of the *cysI* mutant on sulfate in aerobic selective media. The growth rescue required all three factors as well as IPTG induction of the expression plasmids. This result established that the ferredoxin and sulfite reductase components of our pathway were functional and insulated from native metabolism under aerobic conditions.

However, we found that heterologous *zm*FNR was not required to rescue growth when the selective strain was grown without ambient O_2 (Figure 3.7A). While *so*FD and *zm*SIR expression were both still essential, this undesirable background growth indicated that the native metabolic machinery could donate electrons to ferredoxin in anoxic conditions. We did not observe anaerobic growth upon expression of *zm*SIR alone, indicating that electrons were entering the pathway through *so*FD.

Anaerobiosis effects global physiological adaptations in *E. coli* (Iuchi & Weiner 1996). The transition from respiratory to fermentative growth is accompanied by a drop in cytosolic redox potential concomitant with an excess of reducing equivalents generated by glycolysis. A variety of metabolic pathways are expressed specifically in anaerobic conditions to dispose of electrons through electron acceptors including acetate, fumarate, nitrate and hydrogen. Sulfite in our system

represents a high potential electron sink, consistent with the increased tendency for electrons to enter the pathway under these conditions.

To produce the strictest possible connection between the heterologous pathway and strain fitness under selection, we sought to identify and eliminate any endogenous anaerobic electron sources. We selected 6 candidate interacting genes as potential nonspecific electron donors. Candidates were identified based on their homology to known ferredoxin-reducing proteins, with preference given to genes known to be induced anaerobically.

The following six candidate genes were deleted in the BL21(DE3) parent strain: *fpr*, *ydbK*, *hcr*, *yeaX*, *hcaD*, *frdB*. The genes were deleted serially, in a single host strain, in order to expose and eliminate potentially redundant or epistatic interactions of the candidate genes with our pathway (Figure3.7B). The *fpr* gene encodes an anaerobic flavodoxin-NADPH reductase (Jenkins & Waterman 1994). Overexpression of *ydbK*, a putative pyruvate:flavodoxin oxidoreductase, has been shown to drive hydrogenase activity through ferredoxin in *E. coli* (Kalim Akhtar & Jones 2009). The *hcr* locus encodes an anaerobically expressed NADH oxidoreductase that catalyzes the reduction of the hybrid cluster protein Hcp, an iron-sulfur protein with some homology to ferredoxin (Boxhammer et al. 2008). YeaX is a predicted oxidoreductase bearing Fe-S clusters that may associate with the ferredoxin-like YeaW. HcaD encodes a Ferredoxin:NADH reductase involved in the degradation of 3-phenylpropionate. FrdB is an Fe-S protein involved in the anaerobic reduction of fumarate as a terminal electron acceptor. Within the 4-subunit menaquinol-fumarate oxidoreductase complex, FrdB shuttles electrons from the quinone pool to the catalytic flavoprotein FrdA (Cecchini et al. 2002).

We further deleted catalytic subunits of the three characterized endogenous hydrogenases, *hycE*, *hyaB* and *hybC* (Maeda et al. 2007), and the putative but normally silent hydrogenase *hyfG* (Self et al. 2004). All native *E. coli* hydrogenases are of the NiFe-class, unrelated to the FeFe-class and

therefore unlikely to interact directly with our pathway. However, eliminating all native hydrogenases ensures that any H_2 production or consumption in our system could be attributed to the exogenous hydrogenase.

The knockout strains were transformed with *zm*SIR and *so*FD. Strains also received *zm*FNR as a synthetic electron source or an empty vector control. As described in the methods, growth was assayed anaerobically overnight in selective liquid media. The results of these experiments are shown in Figure 3.7.

The sequential knockout of candidate ferredoxin-interacting genes from the selection host improved the insulation of the test pathway from the endogenous redox pool. The first deletion, *fpr,* produced no measurable effect on strain growth. The largest contribution to the elimination of background growth came from the knockout of *ydbK*, with further deletions only modestly decreasing the background growth. These results are consistent with the effects of individual genes observed by Agapakis and colleagues (Agapakis et al. 2010). The mutations had no individual or cumulative deleterious effect on strain growth when *zm*FNR was expressed. We designate as EW11 the final BL21(DE3)-derived strain, which bears the following complete genotype: *E. coli* B F^- *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3) *cysI fpr ydbK hcr yeaX hcaD frdB hycE hyaB hybC hyfG*. All subsequent experiments were performed in EW11 cells.

Sulfide production by our pathway was confirmed spectrophotometrically by the formation of methylene blue, as described in the methods. Wild-type BL21 cells, cultured anaerobically in defined medium, produce small amounts of sulfide: 10 μM (±10 at 95% confidence). No sulfide was detected in cultures of EW11 host cells, consistent with the deletion of *cysI*. Expression of *so*FD and *zm*SIR in EW11 resulted in sulfide accumulation only to a mean level of 2 μM (±1). When *zm*FNR was also expressed as a source of electrons, sulfide levels increased dramatically to 200 μM (±30). Similarly, when the *ca*HydA hydrogenase and maturation factors were expressed, sulfide levels

reached 100 μM (±27). Supplying the hydrogenase with atmospheric H_2 further raised sulfide production to 160 μM (±17). These results are consistent with the design of our pathway as a synthetic source of essential reduced sulfur.

Biochemical hydrogenase O_2-tolerance in situ

We sought to initiate our selection with a wild-type hydrogenase with the highest possible native activity and O_2-tolerance. We reasoned this would improve the probability of evolving an enzyme with properties exceeding those described in nature. This also would allow us to perform our selection in the presence of some O_2, reducing the observed anaerobic background growth. Biochemical techniques allow the *in vitro* determination of purified hydrogenase activity and O_2-tolerance (Baffert et al. 2008). But because genetic selection can be performed only *in vivo*, we assayed hydrogenase O_2-tolerance in cell lysates that approximate the cytosolic context.

E. coli expressing a hydrogenase derived from either *Clostridium acetobutylicum* (*ca*HydA), *Clostridium saccharobutylicum* (*cs*HydA), or *Chlamydomonas reinhardtii* (*cr*HydA), were grown to saturation in liquid culture under strict anaerobiosis. In each case the hydrogenase was coexpressed with the requisite maturation factors HydEF and HydG from *C. reinhardtii*, which are known to mature clostridial FeFe-hydrogenases (Böck et al. 2006). Culture lysates were exposed to O_2 for fixed periods of time and remaining hydrogenase activity was measured biochemically, as described in the methods.

The three hydrogenases were found to differ in both O_2-tolerance and maximal activity levels, as shown in Figure 3.8. The anaerobic activities of the clostridial hydrogenases, *ca*HydA and *cs*HydA, were comparable to each other and both substantially higher than the activity of the *Chlamydomonas* enzyme, *cr*HydA. While our assay controlled for cell density, the *in situ* context of our system did not account for possible differences in expression level, maturation or folding efficiency. Replacing the *C. reinhardtii* maturation factors with those derived from *C. acetobutylicum* yielded the

same *in situ* activity for all three hydrogenases (not shown). While the *in situ* assay does not reflect biochemical specific activities, it measures the effective activity in *E. coli* expressing each hydrogenase, the relevant parameter for our genetic selection.

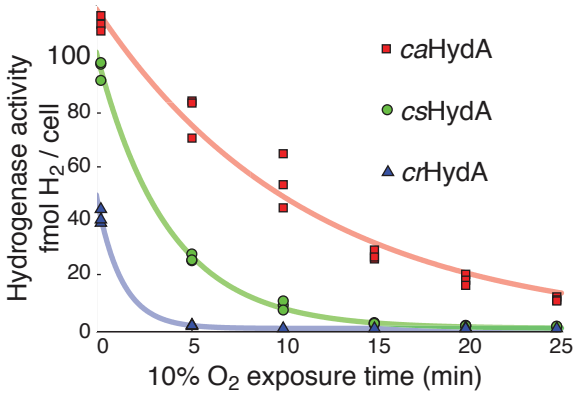


Figure 3.8 O_2-tolerance for three hydrogenases in situ. Cells expressing a hydrogenase from either *C. acetobutylicum* (*ca*HydA), *C. saccharobutylicum* (*cs*HydA) or *C. reinhardtii* (*cr*HydA) were exposed to 10% O_2 at 1 atm total pressure for the indicated times. Remaining activity was assessed with methyl viologen, as described in the methods. Three biological replicates are plotted for each time point. Significant differences in both maximal activity and O_2-tolerance were evident. Hydrogenase inactivation by O_2 was well described by first order kinetics, and the best-fit exponential decay curves are shown. The *ca*HydA enzyme exhibited a characteristic half-life of 8 (±0.8) minutes, the *cs*HydA enzyme 2.7 (±0.2) minutes and *cr*HydA 1.0 (±0.3) minutes, including 95% confidence intervals.

We also observed substantial variation in the natural O_2-tolerance of the three enzymes. In each case the inactivation by O_2 could be well-described by first-order reaction kinetics, resulting in an exponential decrease of activity with time. Exposure to 0.1 atm O_2 partial pressure degraded activity of *ca*HydA with a characteristic half-life of 8 (±0.8) minutes, including a 95% confidence interval. The *cs*HydA enzyme showed a half-life of 2.7 (±0.2) minutes in O_2, and *cr*HydA activity degraded still more rapidly, with a half-life of 1.0 (±0.3) minutes. The half-life measurements are intensive biochemical properties of the enzymes *in situ*, independent of possible differences in hydrogenase expression or maturation levels.

The differences in O_2-tolerance are striking. The clostridial enzymes *ca*HydA and *cs*HydA share 81% amino acid sequence similarity and nearly identical domain architechture. *C. acetobutylicum* and *C. saccharobutylicum* also inhabit similar strictly anaerobic ecological niches (Keis et al. 2001). The algal *cr*HydA is more divergent, only 53% similar to *ca*HydA, yet shares the conserved catalytic domain. *Chlamydomonas*, a eukaryote, exhibits a generally aerobic metabolism. All three enzymes in this experiment receive identically assembled FeFe-cluster active sites from shared maturation factors. Yet the half-life of *ca*HydA in O_2 is twice that of *cs*HydA and 8 times that of *cr*HydA. Because *ca*HydA and *cs*HydA showed higher activity and superior O_2-tolerance, we chose to focus on those enzymes for further study.

The ferredoxin-sulfite reductase interaction

Optimal function of the second step in our pathway requires a robust and specific redox exchange of the mediating Fd with *zm*SIR. We therefore used *zm*FNR as an independent source of electrons to characterize the interaction between our panel of ferredoxins and the *zm*SIR. Growth was measured in anaerobic selective media for strain EW11 with and without *zm*FNR expression. The interactions with both *zm*FNR and *zm*SIR contribute to the ability of a given Fd to facilitate growth in this system. As above, growth in the absence of *zm*FNR expression reflects the tendency of an Fd to receive electrons nonspecifically from the endogenous redox pool.

The results of growth assays from the second test pathway are shown in Figure 3.9. In contrast to the results of the H_2-production assay, we found that the clostridial ferredoxin *ca*FD showed the worst performance in the SIR assay, producing no significant growth. Each of the other ferredoxins tested, *so*FD, *zm*FD and *cr*FD, were able to effectively rescue the sulfide auxotrophy. In the case of the *zm*FD and *cr*FD, we also observed significant background growth in the absence of *zm*FNR expression, indicating nonspecific interactions of these ferredoxins with native metabolism. The *so*FD produced no observable background growth under these conditions.

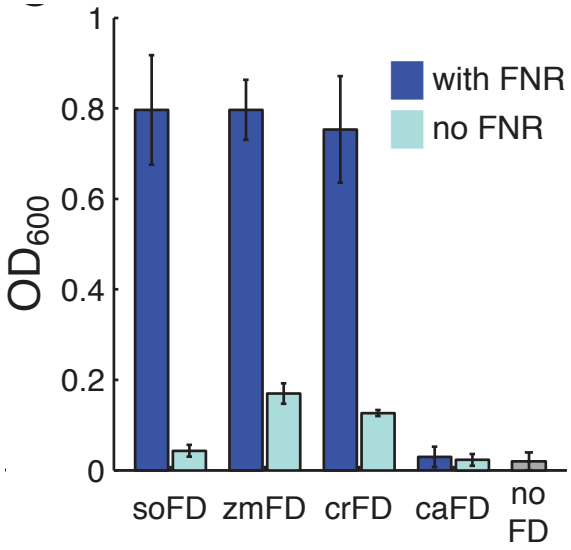


Figure 3.9 FNR-rescued growth with a panel of ferredoxins under anaerobic conditions. Growth with FNR confirms the ability of a ferredoxin to couple with SIR. FNR-independent growth suggests that a ferredoxin is not well insulated from endogenous redox sources. All experiments were conducted under anaerobic conditions with the EW11 strain, as described in the methods. Error bars are 95% confidence intervals.

None of the Fd tested performed optimally in both the SIR and hydrogenase interaction assays (Agapakis et al. 2010). In both cases the differences in performance may be attributed to the distinction between plant-type and bacterial-type ferredoxins. The *so*FD, *zm*FD and *cr*FD proteins all belong to the plant-type class of ferredoxins. These proteins carry electrons in a characteristic Fe_2S_2 active center (Fukuyama 2004). In contrast, *ca*FD is a bacterial-type ferredoxin with a Fe_4S_4 cluster (Bertini et al. 1995). Plant and bacterial ferredoxins share structural and sequence homology (Otaka & Ooi 1989; Matsubara et al. 1979) and can functionally substitute for one another in some cases(Tagawa & Arnon 1962). However, it seems likely that the relative divergence of bacterial *ca*FD precludes an interaction with either *zm*FNR or *zm*SIR, both of which natively pair with plant-type ferredoxins.

We also noted an apparent tendency to higher background activities for ferredoxins with higher redox potentials. Reactions with electrons from the endogenous redox pool may become more thermodynamically favorable as the redox potential of the ferredoxin increases. Among the

plant-type ferredoxins the *so*FD, with a redox potential of -420 mV (Yonekura-Sakakibara et al. 2000), showed the lowest nonspecific activity in both assays. The *zm*FD at -345 mV(Yonekura-Sakakibara et al. 2000) and *cr*FD at -390 mV (Terauchi et al. 2009), exhibited higher backgrounds. This is also consistent with the theory that ferredoxin interactions are governed by Fe-S cluster redox potentials(Moulis & Davasse 1995; Guerrini et al. 2008).

We chose *so*FD as a best functional compromise for the design requirements of our pathway. This ferredoxin clearly demonstrates an ability to function in both redox steps (Figures 3.2 and 3.9), and showed the least nonspecific activity in each. It is also one of the biochemically best characterized ferredoxins and a model for ferredoxin-photosystem interactions (Kovalenko et al. 2011).

O_2 sensitivity of hydrogenase-rescued E. coli

We next characterized *in situ* the behavior of the complete synthetic rescue pathway incorporating *so*FD, *zm*SIR and a hydrogenase electron source. In particular, we sought to quantify the effect of O_2 on the growth of the selection strain in conditions as similar as possible to those that would be encountered in a genetic selection. Strain EW11 expressing *so*FD, *zm*SIR, and hydrogenase maturation factors was transformed with either *ca*HydA, *cs*HydA, or *cr*HydA. Cells were plated at low density on selective plates under custom atmospheres with varying O_2 pressures. Growth was quantified by measuring the size of colonies formed after three days, as described in the methods.

The dose-response relationship of O_2 with various selection strains is depicted in Figure 3.10. Negative control strains expressing only *so*FD, *zm*SIR, and maturation factors showed no measurable growth under any atmosphere (Figure 3.10A). Positive control strains expressing the O_2-tolerant *zm*FNR as an electron source showed robust growth under all conditions. Growth for the

positive control tended to increase with increasing O_2, consistent with the energetic advantages of aerobic metabolism in *E. coli*.

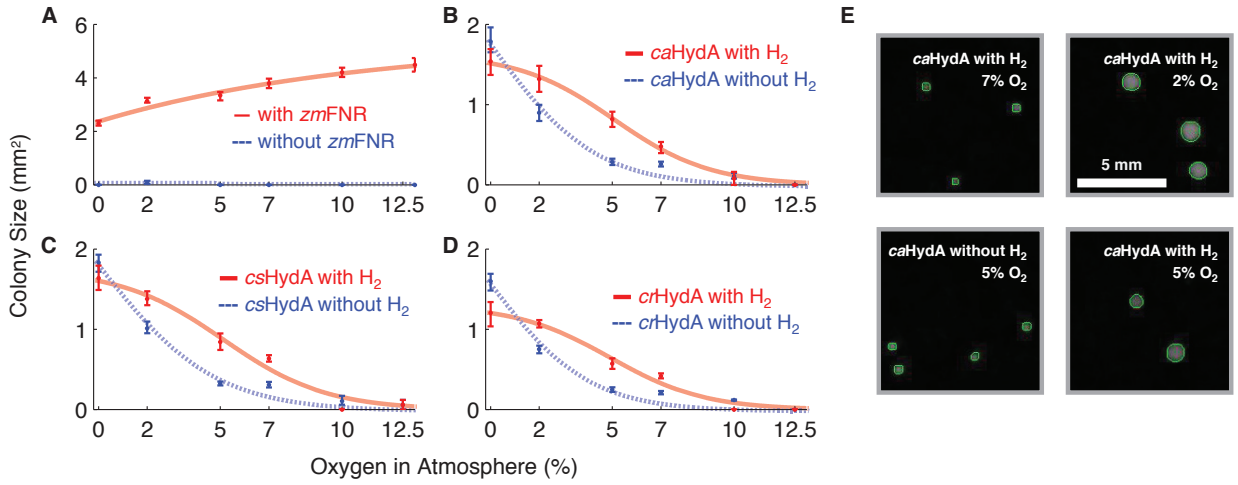


Figure 3.10 Hydrogenase-rescued growth is O_2-dependent and H_2-dependent. The genetically insulated EW11 strain expressing *so*FD, *zm*SIR, and hydrogenase maturation factors was transformed with *zm*FNR (A), empty vector (A), *ca*HydA (B), *cs*HydA (C), or *cr*HydA (D). Cells were plated at low density under custom atmospheres at varying O_2 levels with and without H_2. Colony sizes were measured after 3 days by imaging and automated analysis, as described in the methods. Representative images illustrating the effects of O_2 and H_2 on colony size are displayed in (E), with green circles indicating computationally identified colonies. Growth of strains rescued by each hydrogenase decreased monotonically with increasing O_2 levels, becoming nearly undetectable at 10% O_2. Atmospheric H_2 improved growth but was not required for growth, suggesting an alternative source of reducing equivalents for the hydrogenase. Indicated for each curve is a best-fit sigmoid of the form area = $a/(1+\exp(b\cdot(O_2-c)))$. Error bars are 95% confidence intervals.

Hydrogenase-rescued strains showed O_2-dependent growth (Figure 3.10B-D). For each hydrogenase, growth was the best in 0% O_2 and decreased monotonically until almost no growth was detectable in 10% O_2. Growth supported by a hydrogenase was always less than that observed with *zm*FNR, with colonies less than half as large forming even under strict anaerobiosis. The *ca*HydA enzyme was the most effective hydrogenase at supporting growth over all conditions, followed by *cs*HydA and *cr*HydA. This ordering is consistent with the *in situ* biochemical properties we determined previously. Interestingly, we detected only weak differences in the O_2-dependent growth profiles of the hydrogenases relative to the differences in their *in vivo* activity levels and O_2-tolerance. For example, *ca*HydA shows roughly two-fold higher activity and an eight-fold longer

half-life than *cr*HydA. Yet the strain rescued with *cr*HydA produced colonies only about 30% smaller and reached 50% growth inhibition at the same O_2 level.

Although growth in this assay was strictly dependent on hydrogenase expression, we found that growth did not require the addition of H_2 to the atmosphere. Larger colonies were produced in the presence of H_2 with all hydrogenases at most O_2 levels, yet significant growth was still observed upon replacement of H_2 with N_2. Representative images illustrating the effect of H_2 and O_2 on colony size are shown in Figure 3.10E. H_2-independent growth was more O_2-sensitive, reaching 50% inhibition at roughly 2% ambient O_2 for all hydrogenases, against 5% for H_2-supported growth. When supplied with H_2, strains expressing respectively *ca*HydA, *cs*HydA and *cr*HydA were half-maximally inhibited by O_2 levels of 5.2% (±0.5), 5.7% (±0.4), and 6.1% (±0.7), including 95% confidence intervals. Without H_2, the same level of inhibition was respectively reached at O_2 levels of 2.5% (±0.1), 2.2% (±0.2), and 2.3% (±0.3). This growth could not be attributed to endogenous H_2 production, as the EW11 strain lacks all native hydrogenase activity.

Ferredoxin choice strongly affected the growth of the selection host in the growth-response assay. The substitution of *zm*FD for *so*FD in this pathway resulted in much larger colony sizes by area (data not shown). Yet cells expressing *zm*FD also showed significant nonspecific background growth in the absence of *zm*FNR or *ca*HydA, consistent with results shown in Figure 3.9. We chose to pursue a genetic selection only in strains demonstrating strictly hydrogenase-dependent growth.[7]

DISCUSSION AND CONCLUSIONS

Insulation of a synthetic hydrogen metabolism circuit in bacteria[8]

[7] Barstow et al 2011 further details the results of the SIR-based selection, which I was less directly involved in. The principle findings are summarized in the Discussion and Conclusions section of this chapter.

[8] This section is adapted from Agapakis et al, 2010.

We created an artificial pathway to produce the biofuel hydrogen in *E. coli*. The pathway consists of the proteins pyruvate-ferredoxin oxidoreductase (PFOR), Fd, and a hydrogenase (expressed in the presence of hydrogenase maturation factors). This pathway produces a theoretical maximum of two molecules of hydrogen per input glucose, and still allows acetyl-CoA production from pyruvate. We characterized the relative efficacy of hydrogen production using various combinations of PFOR, Fd, and hydrogenase molecules from different species, and found that PFOR from *D. africanus* in combination with Fd and hydrogenase from *C. acetobutylicum* was the most active pathway, predicted in part by previous *in vitro* data (King et al. 2006; Moulis & Davasse 1995; Pieulle et al. 1999).

To direct electron flow from Fd into hydrogenase, we first deleted genes encoding six other proteins with which PFOR and/or Fd might interact. Of these, only deletion of *ydbK*, encoding a putative *E. coli* PFOR, resulted in enhanced hydrogen production. In addition, in the absence of the PFOR from *D. africanus*, deletion of *ydbK* resulted in a decrease in the background level of hydrogen. These results provide further evidence that *ydbK* is a functional PFOR that can interact with a variety of electron acceptors, particularly the spinach Fd (Kalim Akhtar & Jones 2009).

The iron-sulfur proteins in our synthetic circuit present a modular system, with proteins from disparate species able to interact and produce high levels of hydrogen. Such modular systems are valuable for further synthetic biological manipulation and experimentation. The synthetic pathway presented here is a relatively simple method for the analysis of activities and electron transfer properties of hydrogenases, ferredoxins, and PFOR genes from any number of species, or engineered synthetic electron transfer proteins. These *in vivo* data are a valuable complement to *in vitro* binding constants and kinetic parameters of the enzymes and will be useful in further designing and optimizing microbial systems for hydrogen production.

Such synthetic biological systems can also be used to better understand biological electron transfer systems. The role of ferredoxins in *E. coli* metabolism is poorly characterized, with Fd performing many unknown but required functions in the cell. Here we tested deletions of six iron-sulfur proteins expected to interact with Fd, many of which are previously uncharacterized. While only one gene deletion (Δ*ydbK*) affected our specific hydrogen production pathway, combinatorial deletions may affect hydrogen production in different ways, or may affect other synthetic electron transfer pathways. Further deletions of iron-sulfur oxidoreductases and combinations thereof may lead to a more complete understanding of electron transfer systems in the *E. coli* cytoplasm, as well as the development of a host strain for expression of heterologous electron transfer pathways for synthetic biology. Such a strain would have to retain the ability to mature iron-sulfur clusters but limit the function of proteins that can interact with Fd and Fd oxidoreductases to ensure optimal electron flux through the synthetic pathway. Such specialized strains of *E. coli* may be optimized for other types of synthetic pathway designs and may be better equipped for industrial purposes than proposed "minimal" cells (Forster & Church 2006), as they would retain many of the mechanisms that allow for robust growth and protein expression.

Electron transfer systems such as our hydrogenase pathway are an untapped resource for synthetic biology, which seeks to design biological pathways as predictably as electronic circuits (Andrianantoandro et al. 2006). Electrons are unique metabolites whose movement in biological systems occurs by quantum-mechanical tunneling between protein-bound cofactors such as iron-sulfur clusters. As a result, escape by diffusion into an aqueous phase is avoided, offering distinctive opportunities for control. The circuit described here moves electrons from higher to lower energy, while performing work in the form of hydrogen production. The rationally constructed insulation of the pathway through elimination of side reactions, interaction surface optimization, and protein

fusion or scaffolding[9], indicate that all four methods are viable for synthetic circuit design and all strategies may play a role in the evolution of complex isolated circuits in natural metabolism. This type of synthetic-biological analysis may yield insights into natural mechanisms for controlling electron flow, and may provide new approaches for metabolic engineering and bioenergy.

Selection systems for the directed evolution of oxygen-tolerant [FeFe]-hydrogenases

We describe two genetic selection circuits that potentially couple *E. coli* growth to hydrogenase function. One strategy, based on the hydrogen producing PFOR-Fd-hydrogenase interaction, did not yield a strain suitable for selection experiments. Unlike the SIR-based selection strategy, this selection required the hydrogenase-dependent circuit to integrate into central metabolism; it is possible that the amount of pathway throughput we observed (indirectly via hydrogen production *in vivo*) was insufficient to support growth. In all experiments described in this chapter, oxidoreductases, Fd, HydA, and the maturation factors were all overexpressed via T7 polymerase. This protein overexpression exacerbated the cellular metabolic burden. Ultimately, the superior performance of the SIR-based selection, which supplied essential reduced sulfur but was not involved in central carbon metabolism, led us to choose this system for oxygen tolerance experiments. Even the SIR-based selection, however, failed to yield hydrogenases with improved oxygen tolerance (Barstow et al. 2011)[10].

Outcome of the SIR-based selection[11]

[9] See Agapakis et al 2010 for details on the interaction surface optimization and protein fusion/scaffolding approaches

[10] I am omitting sections of the Barstow et al 2011 manuscript that cover the generation of mutated hydrogenases and the oxygen tolerance selections.

[11] This paragraph and the following paragraphs in this section are adapted from Barstow et al 2011. Although I did not conduct experiments related to the generation and selection of mutated

The genetic selection for O_2-tolerance produced no hydrogenases with O_2-tolerance exceeding that of the wild-type, and only one with a comparable activity level. Instead, we observed a widespread loss of H_2-production activity among selected hydrogenases (Barstow et al. 2011). We speculate that the mutants identified by genetic selection confer growth by enhancing H_2-independent electron transfer to ferredoxin. Consistent with this hypothesis are the results of Figure 3.10, indicating that hydrogenase, but not H_2, is essential for growth under selection.

We found significant enrichment of mutations neutralizing positive surface charges of the hydrogenase. Electrostatic forces are known to have an important role in the kinetics and specificity of intermolecular interactions (Mittag et al. 2010; Davis & McCammon 1990). Fd proteins such as those used in our pathway display numerous and conserved negatively charged surface residues, which are thought to govern the specific recognition of various Fd redox partners (Moulis & Davasse 1995). Site-directed mutagenesis of lysine residues on the surface of *Anabaena* FNR was found to block interaction with its native Fd (Schmitz et al. 1998). Our selection pathway employs a non-native pairing of spinach Fd and clostridial hydrogenase, invoking suboptimally co-adapted electrostatic interactions. The mutation of surface lysines may enhance ferredoxin-hydrogenase charge complementarity (Chang et al. 2007). This could allow for more efficient electron transfer to ferredoxin, an essential activity for host viability.

Our selection also isolated a number of highly truncated hydrogenase variants which were found to retain some function. Variants BB22 and BB05 lacked portions of the small C-terminal hydrogenase subunit but were still competent for hydrogen production (Barstow et al. 2011). Similarly truncated putative hydrogenases have been identified in the termite hindgut metagenome (Warnecke et al. 2007). Even more severely truncated variants showed no hydrogenase activity, but

---

hydrogenases, I include discussion of the outcome of these experiments to place this project in the broader context of biohydrogen production.

were nevertheless capable of rescuing growth. Variant BB09, for example, retained only the N-terminal ferredoxin-like domains, which were nevertheless sufficient to facilitate electron transfer to our synthetic pathway. Future work to identify the minimal structural elements required for hydrogenase function may help to structurally couple hydrogenases to other electron sources and sinks.

Our pathway allows O_2-tolerant electron sources such as *zm*FNR to be distinguished from O_2-sensitive sources such as *ca*HydA by their effects on host fitness. Therefore an O_2-tolerant hydrogenase, once produced, could in principle be isolated using our selection. That no such hydrogenase was found suggests that other mutations exist to alter hydrogenase redox activity independently of the described O_2-sensitive catalytic core (Stripp et al. 2009). Mutations of this sort may be more common than those specifically altering properties of the active site. The evolution of O_2-tolerance may also require more simultaneous mutations than were sampled here, or more extensive structural alterations. The structural features that optimize O_2-tolerance might also change when the direction of hydrogenase activity favors consumption versus production. Future efforts may benefit from combining genetic selection with high-throughput techniques to biochemically characterize hydrogenases, currently in development (Stapleton & Swartz 2010).

The data presented in Figure 3.10 suggest that each hydrogenase can be reduced by an unknown endogenous electron source other than H_2. This electron source is eliminated by oxygen, but apparently through a different mechanism than that which directly inactivates the catalytic H-cluster (Stripp et al. 2009). Such a model would explain the apparent discrepancy between our *in vivo* and *in vitro* O_2-tolerance assays. While Figure 3.7 shows different kinetics for the oxygen inactivation of each hydrogenase *in vitro*, Figure 3.10 shows all three enzymes support comparable O_2-tolerant growth. Hydrogenase activity in our system might also be limited by oxygen sensitivity of the maturation factors, rather than the mature enzymes. To our knowledge, potential interactions of

HydEFG with oxygen have not yet been directly examined. An *in vivo* selection system would become even more valuable in such a case, as mutagenesis and selection could naturally be extended to the maturation factors.

We have shown that engineering can successfully insulate a synthetic electron transfer pathway from the endogenous *E. coli* redox pool. Minimizing losses to the cell through insulation of an artificial pathway allows more rational control of an engineered metabolic flux. We have demonstrated the use of convenient *in vivo* assays to validate isolated components of a synthetic H_2 metabolism. The results of our assays revealed trade-offs in the choice of pathway components, allowing compromises to meet design goals. Finally, we have successfully tied an essential part of cellular metabolism, the synthesis of cysteine, to hydrogenase activity. By eliminating or disabling this activity with O_2, we can halt cellular metabolism. These results demonstrate the utility of this pathway in a genetic selection for O_2-tolerant hydrogenases. We anticipate that future work to characterize the H_2-independent hydrogenase activity, and to optimize ferredoxin-hydrogenase electron transfer, will allow for more strict selection of hydrogen catalysts with desired properties.

MATERIALS AND METHODS

Insulation of a synthetic hydrogen metabolism circuit in bacteria[12]

Plasmids and cloning

All cloning was done in *E. coli* DH5α. Hydrogenase genes from *Chlamydomonas reinhardtii* and the ferredoxin I gene from *Spinacia olearcea* were commercially synthesized by Codon Devices (Cambridge, MA), codon optimized for expression in *Saccharomyces cerevisiae* and acceptable for use in *E. coli* for wide applicability (see Appendix A, figure S1 for nucleotide sequences). Hydrogenases from *Clostridium acetobutylicum* and *Clostridium saccharobutylicum* were cloned from plasmids received

[12] This section is adapted from Agapakis et al, 2010.

from Matthew Posewitz (National Renewable Energy Laboratory, Golden, CO). Hydrogenase genes HydA and HydB were cloned from *Shewanella oneidensis* using colony PCR of bacterial cultures from Colleen Hansel (Harvard University, Cambridge, MA). *Thermotoga maritima* HydA was cloned from genomic DNA provided by Kenneth Noll (University of Connecticut, Storrs, CT). PFOR and ferredoxin [27] were cloned from *Clostridium acetobutylicum* genomic DNA (ATCC, Manassas, VA). *Zea mays* ferredoxin was cloned from genomic DNA isolated from corn using DNeasy Plant Mini Kit (Qiagen, Valencia, CA). PFOR from *Desulfovibrio africanus* was isolated from plasmid pLP1 [28] provided by Laetitia Pieulle (Centre National de la Recherche Scientifique, Marseille, France) and *ydbK* was obtained through colony PCR of *E. coli* BL21.

Protein expression

All protein expression and hydrogenase activity assays were performed in *E. coli* BL21 (DE3). Cells were transformed with modified pCDF-duet with *C. reinhardtii* HydEF in MCS1 and *C. reinhardtii* HydG in MCS2, and with modified pACYC-duet with *C. acetobutylicum* PFOR or *E. coli ydbK* in MCS1 or *Desulfovibrio africanus* PFOR cloned into the downstream NdeI and AvrII sites of MCS2. Hydrogenase/ferredoxin pairs were transformed either in each multiple cloning site of modified pET-duet, or for the *S. oneidensis* hydrogenase HydA in MCS1, HydB in MCS2, and ferredoxin in MCS1 of modified pCOLA-duet. To compare *in vitro* hydrogen production using maturation factors from *Clostridium acetobutilicum,* we used plasmids provided by Matthew Posewitz (pCDF-duet with CaHydE in MCS1, CaHydF in MCS2 and pET-duet with CaHydA in MCS1 and CaHydG in MCS2 (King et al. 2006)).

***E. coli* Gene Deletion**

Hydrogenase knockout (Δ*hycE,* Δ*hyaB,* Δ*hybC*) and Δ*fpr,* Δ*ydbK,* Δ*hcr,* Δ*yeaX,* Δ*hcaD,* or Δ*frdB* single deletion strains were made by sequential P1 transduction from the Keio collection (Baba et al. 2006) into BL21(DE3) Δ*tonA*, followed by removal of the Kan^R marker by standard procedures.

SDS-Page and Western Blotting

E. coli cells were lysed with Bacterial Protein Extraction Reagent (B-PER, Pierce, Rockford, IL), protein samples were normalized using the Bradford assay (Bio-Rad, Hercules, CA), diluted into SDS-PAGE loading buffer and loaded onto a 4-20% Tris/glycine/SDS acrylamide gel. a-Strep-tag II antibody (HRP-conjugated, Novagen, Gibbstown, NJ) or a-ferredoxin primary antibody (Agrisera, Vännäs, Sweden) and a-Rabbit IgG secondary antibody were used.

Hydrogen production assays

Bacterial cultures were grown aerobically for two hours until reaching an OD_{600} of approximately 0.15 in LB media with appropriate antibiotic (50 μg/ml ampicillin, 25 μg/ml spectinomycin, 25 μg/ml kanamycin, and/or 12.5 μg/ml chloramphenicol) in 40 ml glass serum vials, induced with 1mM IPTG (and 2 μg/ml anhydrous tetracycline when relevant for the induction of scaffold proteins) and sparged with argon. For the methyl viologen assay, adapted from King et. al. (King et al. 2006), vials were sparged for 2 hours and then lysed with assay buffer containing 50 mM Tris pH 7.0, 50% B-PER, 10mM methyl viologen (Sigma, St. Louis, MO) and 50mM sodium dithionite (Fisher, Pittsburgh, PA), the vials were capped with rubber septa, the cells were vortexed and allowed to rock overnight at room temperature. Hydrogen concentration in the headspace gas was measured by gas chromatography (Shimadzu GC-14A). *In vivo* hydrogen production assays were performed in a similar fashion, except that cultures were supplemented with 0.5% glucose at the time of IPTG induction, sparged for 30 minutes and simply capped and shaken ovenight at 37°C before measuring headspace gas composition. Glucose curves were measured in cells pretreated overnight with 1mM IPTG then immediately diluted into LB + variable glucose + 1mM IPTG, then sparged and grown overnight. All hydrogen production values were normalized to an OD_{600} of 0.15.

A synthetic system links FeFe-hydrogenases to essential E. coli sulfur metabolism[13]

Cloning and gene synthesis

All cloning was performed in *E. coli* DH5α using standard BioBrick assembly techniques(Phillips & Silver 2006). Final constructs were assembled in commercial Duet vectors (Novagen) with multiple cloning sites modified to accept BioBrick parts. The plasmids used in these experiments are listed in table 1. Complete vector sequences are provided as supplementary information.

Zea mays sulfite reductase (GenBank BAA23641) was cloned from total RNA. The ferredoxin-NADPH reductase (FNR) gene from *Zea mays* (GenBank AAB40034) was synthesized by Codon Devices. Chloroplast transit peptides were omitted from all plant-derived constructs. The codon usage of synthetic genes was optimized by the manufacturer for heterologous expression.

Selective and induction media

Selective media was a standard M9 formulation, supplemented with additional glucose, sulfate, ferric iron and a rich mix of supplements less cysteine and methionine. Induction media for hydrogenase expression was LB with added glucose, ferric iron, phosphate buffer and Baker's antifoam reagent. Exact media recipes are provided as additional files.

Anaerobic technique and custom atmospheres

Anaerobic liquid culture was performed in 40ml serum vials sparged with nitrogen and sealed with SubaSeal® rubber septa (Sigma-Aldrich). To maintain anaerobiosis during handling, samples were drawn and reagents added by piercing the septa with non-coring syringe needles.

Agar plates were incubated under defined gas mixtures within sealed Vacu-Quick jars (Almore International). The ambient atmosphere was removed by several cycles of evacuation and

[13] This section is adapted from Barstow et al 2011. Materials and methods redundant with those in Agapakis et al 2010 have been removed.

replacement with pure nitrogen before supplying a custom atmosphere. Aluminosilicate desiccant packets were added to prevent moisture accumulation within the jars.

Growth assays

Cells were grown to saturation in induction media and washed 3x with phosphate-buffered saline (PBS). Cells were resuspended in selective media at an initial OD_{600} of 0.01. Final ODs were measured after 18 hours of growth at 37°C. Anaerobic conditions, when appropriate, were introduced as described above.

***In situ* hydrogenase activity assays**

Cells were grown to saturation in anaerobic induction media. Samples were drawn to determine cell density by OD. Fresh serum vials containing 25 ml of induction LB were anaerobically inoculated with 10^8 cells (~5ml). Following incubation at 37°C for 2 hours, hydrogen production was stopped by the addition 2ml methanol. Accumulated headspace hydrogen was measured by gas chromatography (Shimadzu GC-14A).

***In vitro* hydrogenase activity assays**

Hydrogenase activities were measured with a biochemical methyl viologen assay adapted from King *et al.*(King et al. 2006). Hydrogenase-expressing *E. coli* were grown to saturation in 20 mL of induction media under anaerobic conditions. Samples were drawn to determine cell density by OD. Cells were lysed with 1ml of lysis buffer consisting of 20ml B-PER II protein extraction reagent (Thermo scientific), 500 μL Baker's antifoam B, 100 units DNAse I and 50mg dithionite. Antifoam and DNAse I were added to prevent foaming of the lysate. Dithionite served to scavenge dissolved O_2 in the buffer. Lysis continued for 15 minutes under continuous nitrogen sparging. Lysed cultures were sealed and injected with 1 mL of methyl viologen assay buffer consisting of 20 mL 1 M Tris at pH 8, 300 mg methyl viologen and 3 g dithionite. Following 2 hours of incubation at

37 °C, H_2 production was stopped with 2 mL of methanol. Accumulated headspace hydrogen was measured by gas chromatography.

Sulfide production assays

Sulfide production was measured spectrophotometrically by the methylene blue method(Cline 1969). Bacteria were grown to saturation in selective media with IPTG and appropriate antibiotics, supplemented with 100 mg/L cysteine to allow growth of all strains. Cultures were diluted 1:50 to 25 ml in 40 ml sealed serum vials. Vials were flushed with either pure N_2 or 40% H_2 in N_2 and grown for 6 hours at 37 °C. Cuvettes were prepared with 2.5 mL assay solution containing N,N-Dimethyl-1,4-phenylenediammonium dichloride (200 μM), FeCl (600 μM) and HCl (60mM). 25 μL of filtered media was added to assay cuvettes and allowed to react for 1 hour. Absorbance was measured at 660 nm and compared to a standard curve to calculate sulfide concentrations. Statistics were inferred from three biological replicates.

Hydrogenase O_2-tolerance assays

Hydrogenase half-lives in O_2 were measured using an adaptation of the *in vitro* activity assay above. Following anaerobic lysis, a customized sparging apparatus was used to bubble O_2 through the cultures at a constant partial pressure of 0.1 atm for defined time points between 5 and 25 minutes. Cultures were then flushed for 5 minutes with pure nitrogen before being sealed and assayed with methyl viologen assay buffer, as above. Measurements at each time point were taken for 3 biological replicates.

Colony growth assays under custom atmospheres

Cells were grown to saturation in induction media. Residual nutrients were removed by 3x washing with PBS. Cell densities were measured with a hemocytometer and diluted to a final concentration of 1 cell/μL. Fifty microliters (≈ 50 cells) were dispensed onto selective plates and transferred Vacu-Quick jars. The jars were filled with 15% H_2, O_2 varying from 0-12.5%, and a

balance of nitrogen to a total internal pressure of 1 atm. Incubation was carried out at 37 ˚C for 72 hours. Plates were photographed in an inverted camera stage (Figure 3.10E). Colonies were identified and sized with an image analysis script implemented in Matlab (MathWorks). Each data point represents size data collected from roughly 50 individual colonies.

Design of a PFOR-based genetic selection for oxygen tolerant hydrogenases

Gene cloning and gene deletion were conducted as described above. Hydrogen production *in vivo* was conducted as in (Agapakis et al. 2010). For the growth assays described in Table 3.1, TYA media (Zhu et al. 2008) was used. TYA media contains (per liter): 10.0 g tryptone, 5.0 g NaCl, 1.0 g yeast extract, 1.36 g Na(CH3COO) ·3H2O. Cells were grown to saturation at 37°C in liquid TYA media, then back-diluted to an OD600 of approximately 0.3 into TYA media containing 100 μM IPTG. After 2 days of growth, 50 μL of each culture were transferred to solid agar TYA plates containing 100 μM IPTG. Plates were grown for 72 hours at 37°C under ambient air or a 100% nitrogen atmosphere as described above.

Supporting information for this chapter is available as Appendix A.

ACKNOWLEDGEMENTS

Agapakis et al 2010: We would like to acknowledge David Savage, Devin Burrill, Bruno Afonso, Noah Taylor, and Dominique Frueh for helpful discussion and comments on the manuscript, Craig Blain, Renan Antonio Escalante Chung, and Bennett Caughey for technical assistance, and John Dueber, Matthew Posewitz, Colleen Hansel, Laetitia Pieulle, and Kenneth Noll for providing materials. This material is based upon work supported by the Army Research Office Award No. W911NF-09-1-0226. CMA and EHW are supported by NSF Graduate Research fellowships and PMB is funded by the Harvard University Center for the Environment Graduate Consortium.

Barstow et al 2011: We thank Daniel Ducat and Devin Burrill for manuscript revisions. David Savage, Zeev Waks, and Jeffrey Way and the Silver lab provided useful discussions. Remy Chait and Laura Stone assisted with image acquisition. Bruno Afonso assisted with image analysis. Andrew Bergman aided parallelized computations on the HMS Orchestra Cluster. Jim Horn constructed experimental apparatus. Funding was provided by NSF graduate research fellowships to EHW and CMA, an NSF SynBERC fellowship to PMB, and an NRSA post-doctoral fellowship to BB. Research was supported by Army Research Office Award No. W911NF-09-1-0226.

Competing interests: EHW, PMB, CMA, and PAS are inventors of International Patent Application No. PCT/US2009/058361, "In-Vivo Selection System For ID Iron (FEFE) Hydrogenase Properties."

REFERENCES

Agapakis, C.M. et al., 2010. Insulation of a synthetic hydrogen metabolism circuit in bacteria. *Journal of Biological Engineering*, 4, p.3.

Aguirre de Carcer, I. et al., 2006. Active-site models for iron hydrogenases: reduction chemistry of dinuclear iron complexes. *Inorganic chemistry*, 45(20), pp.8000–8002.

Akhtar, M. & Jones, P., 2008. Deletion of iscR stimulates recombinant clostridial Fe-Fe hydrogenase activity and H(2)-accumulation in Escherichia coli BL21(DE3). *Applied Microbiology and Biotechnology*.

Andrianantoandro, E. et al., 2006. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2, p.2006.0028.

Armstrong, F., 2009. Dynamic electrochemical experiments on hydrogenases. *Photosynthesis Research*.

Baba, T. et al., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, p.2006.0008.

Baffert, C. et al., 2008. Hydrogen-activating enzymes: activity does not correlate with oxygen sensitivity. *Angewandte Chemie (International ed in English)*, 47(11), pp.2052–2054.

Barstow, B. et al., 2011. A synthetic system links FeFe-hydrogenases to essential E. coli sulfur metabolism. *Journal of Biological Engineering*, 5(1), pp.7–7.

Benemann, J.R. et al., 1973. Hydrogen evolution by a chloroplast-ferredoxin-hydrogenase system. *Proceedings of the National Academy of Sciences of the United States of America*, 70(8), pp.2317–2320.

Bertini, I. et al., 1995. Solution structure of the oxidized 2[4Fe-4S] ferredoxin from Clostridium pasteurianum. *European journal of biochemistry / FEBS*, 232(1), pp.192–205.

Boxhammer, S. et al., 2008. Characterization of the recombinant Rieske [2Fe-2S] proteins HcaC and YeaW from E. coli. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine*, 21(4), pp.459–467.

Boxma, B. et al., 2005. An anaerobic mitochondrion that produces hydrogen. *Nature*, 434(7029), pp.74–79.

Böck, A. et al., 2006. Maturation of hydrogenases. *Advances in microbial physiology*, 51, pp.1–71.

Casalot, L. & Rousset, M., 2001. Maturation of the [NiFe] hydrogenases. *Trends in Microbiology*, 9(5), pp.228–237.

Cecchini, G. et al., 2002. Succinate dehydrogenase and fumarate reductase from Escherichia coli. *Biochimica et biophysica acta*, 1553(1-2), pp.140–157.

Chang, C.H. et al., 2007. Atomic resolution modeling of the ferredoxin:[FeFe] hydrogenase complex from Chlamydomonas reinhardtii. *Biophysical journal*, 93(9), pp.3034–3045.

Cline, J.D., 1969. Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnology and Oceanography*, 14(3), pp.454–458.

Davis, M.E. & McCammon, J.A., 1990. Electrostatics in biomolecular structure and dynamics. *Chemical Reviews*, 90(3), pp.509–521.

Demuez, M. et al., 2007. Complete activity profile of Clostridium acetobutylicum [FeFe]-hydrogenase and kinetic parameters for endogenous redox partners. *FEMS Microbiology Letters*, 275(1), pp.113–121.

Feist, A.M. et al., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3, p.121.

Forster, A.C. & Church, G.M., 2006. Towards synthesis of a minimal cell. *Molecular Systems Biology*, 2, p.45.

Fukuyama, K., 2004. Structure and function of plant-type ferredoxins. *Photosynthesis Research*, 81(3), pp.289–301.

Ghirardi, M.L. et al., 2007. Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms. *Annual review of plant biology*, 58, pp.71–91.

Ghirardi, M.L. et al., 2009. Photobiological hydrogen-producing systems. *Chemical Society Reviews*, 38(1), pp.52–61.

Gou, P. et al., 2006. Higher order structure contributes to specific differences in redox potential and electron transfer efficiency of root and leaf ferredoxins. *Biochemistry*, 45(48), pp.14389–14396.

Guerrini, O. et al., 2008. Characterization of two 2[4Fe4S] ferredoxins from Clostridium acetobutylicum. *Current microbiology*, 56(3), pp.261–267.

Ihara, M. et al., 2006. Light-driven hydrogen production by a hybrid complex of a [NiFe]-hydrogenase and the cyanobacterial photosystem I. *Photochemistry and Photobiology*, 82(3), pp.676–682.

Iuchi, S. & Weiner, L., 1996. Cellular and molecular physiology of Escherichia coli in the adaptation to aerobic environments. *Journal of biochemistry*, 120(6), pp.1055–1063.

Jenkins, C.M. & Waterman, M.R., 1994. Flavodoxin and NADPH-flavodoxin reductase from Escherichia coli support bovine cytochrome P450c17 hydroxylase activities. *The Journal of biological chemistry*, 269(44), pp.27401–27408.

Kalim Akhtar, M. & Jones, P., 2009. Construction of a synthetic YdbK-dependent pyruvate:H(2) pathway in escherichia coli BL21(DE3). *Metabolic Engineering*.

Kanehisa, M. et al., 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 36(Database issue), pp.D480–4.

Keis, S., Sullivan, J.T. & Jones, D.T., 2001. Physical and genetic map of the Clostridium

saccharobutylicum (formerly Clostridium acetobutylicum) NCP 262 chromosome. *Microbiology (Reading, England)*, 147(Pt 7), pp.1909–1922.

King, P.W. et al., 2006. Functional studies of [FeFe] hydrogenase maturation in an Escherichia coli biosynthetic system. *Journal of Bacteriology*, 188(6), pp.2163–2172.

Kontur, W.S., Noguera, D.R. & Donohue, T.J., 2011. Maximizing reductant flow into microbial H(2) production. *Current Opinion in Biotechnology*.

Kovalenko, I.B. et al., 2011. Computer simulation of interaction of photosystem 1 with plastocyanin and ferredoxin. *Bio Systems*, 103(2), pp.180–187.

Maeda, T., Sanchez-Torres, V. & Wood, T.K., 2007. Escherichia coli hydrogenase 3 is a reversible enzyme possessing hydrogen uptake and synthesis activities. *Applied Microbiology and Biotechnology*, 76(5), pp.1035–1042.

Matsubara, H. et al., 1979. Structural and evolution of chloroplast- and bacterial-type ferredoxins. *UCLA forum in medical sciences*, (21), pp.245–266.

McCord, J.M., 2000. The evolution of free radicals and oxidative stress. *The American journal of medicine*, 108(8), pp.652–659.

Mittag, T., Kay, L.E. & Forman-Kay, J.D., 2010. Protein dynamics and conformational disorder in molecular recognition. *Journal of molecular recognition : JMR*, 23(2), pp.105–116.

Moulis, J.M. & Davasse, V., 1995. Probing the role of electrostatic forces in the interaction of Clostridium pasteurianum ferredoxin with its redox partners. *Biochemistry*, 34(51), pp.16781–16788.

Mulder, D.W. et al., 2011. Insights into [FeFe]-Hydrogenase Structure, Mechanism, and Maturation. *Structure (London, England : 1993)*, 19(8), pp.1038–1052.

Nakamura, M., Saeki, K. & Takahashi, Y., 1999. Hyperproduction of recombinant ferredoxins in escherichia coli by coexpression of the ORF1-ORF2-iscS-iscU-iscA-hscB-hs cA-fdx-ORF3 gene cluster. *Journal of biochemistry*, 126(1), pp.10–18.

Otaka, E. & Ooi, T., 1989. Examination of protein sequence homologies: V. New perspectives on evolution between bacterial and chloroplast-type ferredoxins inferred from sequence evidence. *Journal of Molecular Evolution*, 29(3), pp.246–254.

Page, C.C. et al., 1999. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature*, 402(6757), pp.47–52.

Phillips, I. & Silver, P., 2006. A new biobrick assembly strategy designed for facile protein engineering. *DSpace. MIT Artificial Intelligence Laboratory: MIT Synthetic Biology Working Group, Massachusetts Institute of Technology, Cambridge, MA. http://hdl. handle. net/1721.1/32535*.

Pieulle, L. et al., 1999. Structural and kinetic studies of the pyruvate-ferredoxin oxidoreductase/ferredoxin complex from Desulfovibrio africanus. *European journal of biochemistry*

/ *FEBS*, 264(2), pp.500–508.

Pieulle, L., Magro, V. & Hatchikian, E.C., 1997. Isolation and analysis of the gene encoding the pyruvate-ferredoxin oxidoreductase of Desulfovibrio africanus, production of the recombinant enzyme in Escherichia coli, and effect of carboxy-terminal deletions on its stability. *Journal of Bacteriology*, 179(18), pp.5684–5692.

Savage, D.F., Way, J. & Silver, P.A., 2008. Defossiling fuel: how synthetic biology can transform biofuel production. *ACS chemical biology*, 3(1), pp.13–16.

Schmitz, S. et al., 1998. Interaction of positively charged amino acid residues of recombinant, cyanobacterial ferredoxin:NADP+ reductase with ferredoxin probed by site directed mutagenesis. *Biochimica et biophysica acta*, 1363(1), pp.85–93.

Self, W.T., Hasona, A. & Shanmugam, K.T., 2004. Expression and regulation of a silent operon, hyf, coding for hydrogenase 4 isoenzyme in Escherichia coli. *Journal of Bacteriology*, 186(2), pp.580–587.

Shima, S. et al., 2008. The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. *Science (New York, NY)*, 321(5888), pp.572–575.

Stapleton, J.A. & Swartz, J.R., 2010. Development of an In Vitro Compartmentalization Screen for High-Throughput Directed Evolution of [FeFe] Hydrogenases. *PLoS ONE*, 5(12), p.e15275.

Stripp, S. et al., 2009. How oxygen attacks [FeFe] hydrogenases from photosynthetic organisms. *Proceedings of the National Academy of Sciences of the United States of America*.

Tagawa, K. & Arnon, D.I., 1962. Ferredoxins as electron carriers in photosynthesis and in the biological production and consumption of hydrogen gas. *Nature*, 195, pp.537–543.

Terauchi, A.M. et al., 2009. Pattern of expression and substrate specificity of chloroplast ferredoxins from Chlamydomonas reinhardtii. *Journal of Biological Chemistry*, 284(38), pp.25867–25878.

Veit, A. et al., 2008. Constructing and testing the thermodynamic limits of synthetic NAD(P)H:H2 pathways. *Microbial Biotechnology*, 1(5), pp.382–394.

Vignais, P.M. & Billoud, B., 2007. Occurrence, classification, and biological function of hydrogenases: an overview. *Chemical Reviews*, 107(10), pp.4206–4272.

Warnecke, F. et al., 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 450(7169), pp.560–565.

Yonekura-Sakakibara, K. et al., 2000. Analysis of reductant supply systems for ferredoxin-dependent sulfite reductase in photosynthetic and nonphotosynthetic organs of maize. *Plant Physiol*, 122(3), pp.887–894.

Zhu, Y. et al., 2008. High glycolytic flux improves pyruvate production by a metabolically engineered Escherichia coli strain. *Applied and Environmental Microbiology*, 74(21), pp.6649–6655.

Chapter 4

Circadian Rhythm Controls Light-Independent Metabolic Oscillations in Cyanobacteria[1]

ABSTRACT

The metabolism of cyanobacteria is critically dependent on light availability. Thus, cyanobacteria have evolved robust circadian clocks to synchronize with this essential resource. Although the global circadian regulation of gene expression is well documented, the circadian control of metabolism has not been measured at the network level. To determine the impact of circadian regulation on metabolism, we conducted a metabolomic analysis of the model cyanobacterium *Synechococcus elongatus* PCC 7942 using targeted mass spectrometry. We detected over 200 metabolites, tracking variation in metabolic abundance during a 24-hr day/night period as well as during 48 hrs under constant light. The majority of metabolites appear to be controlled by light input rather than the circadian clock. However, we also identified a subset of metabolites that oscillate in abundance with a 24-hr period. Surprisingly, these metabolic oscillations are governed by the Kai clock despite constant light availability. Furthermore, circadian-dependent disaccharide oscillations appear to reinforce clock entrainment. Thus, this work provides insight into the extent of transcriptional control over metabolite concentration and into the dynamic regulation of photosynthetic metabolism.

[1] This chapter has been submitted for publication as: Boyle, P.M., Taylor, N.D., Savage, D.F., Asara, J.M., and Silver, P.A. Circadian Rhythm Controls Light-Independent Metabolic Oscillations in Cyanobacteria.
Contributions: P.M.B, D.F.S., and P.A.S. designed research; P.M.B, N.D.T., D.F.S., and J.M.A. performed research; P.M.B. and P.A.S. analyzed data and wrote the paper.

INTRODUCTION

Cells must adjust their metabolism to compensate for changes in their environment. For photoautotrophic organisms, light is a primary resource that is periodically available. As a result, cyanobacteria are the only prokaryotes to have evolved circadian clocks; these clocks allow cyanobacteria to anticipate the diurnal cycling of light intensity and to adjust to seasonal trends in day length (Dong & Golden 2008). In *Synechococcus elongatus* PCC 7942 (hereafter *S. elongatus*), it has been estimated that the expression of 30-100% of genes is governed by the circadian rhythm (Vijayan et al. 2009; Ito et al. 2009). Loss of function mutations in circadian clock genes decrease the fitness of *S. elongatus* under diurnal conditions (Woelfle et al. 2004).

The KaiC protein serves as the master regulator of the circadian clock in *S. elongatus* (Tomita et al. 2005; Nishiwaki et al. 2004). The proteins KaiA and KaiB, along with ATP, are necessary and sufficient to establish oscillations in the phosphorylation state of KaiC *in vitro* (Wold et al. 2001; Nakajima et al. 2005). Light-driven changes in intracellular energy state appear to entrain the Kai clock (Rust et al. 2011), but the influence of circadian regulation on intracellular metabolite concentrations is not well understood. Furthermore, feedback between metabolites other than ATP/ADP and the circadian clock may be important to regulate of circadian rhythms *in vivo*.

Metabolites under circadian control are expected to oscillate in an entrainment dependent and light independent manner. Entrainment is a hallmark feature of circadian clocks: once entrained by diurnal stimuli, circadian clocks continue to oscillate with an approximate 24-hr period even if the stimuli are removed (Roenneberg & Foster 1997). In S. elongatus, growth under a 12hr:12hr light:dark rhythm ("LD") results in 24-hr Kai oscillations for 3 days following a shift to a 24hr:0hr light:dark rhythm ("LL") (Kondo et al. 1993; Golden 2007). Systems that are regulated by the circadian rhythm also oscillate post-entrainment, as has been demonstrated with gene expression in

S. elongatus (Vijayan et al. 2009; Ito et al. 2009). Therefore, we expect metabolites that accumulate as a function of enzyme transcript levels (and by extension a function of the Kai clock) to oscillate in LL conditions post-entrainment.

In this study we tracked over 200 metabolites in entrained cultures of *S. elongatus*. By tracking the abundance of metabolites post-entrainment, we directly observed the extent of circadian control over metabolism. Metabolism in *S. elongatus* depends on both light input and circadian entrainment as regulators of the metabolic network. Since the transcriptome of *S. elongatus* is so tightly controlled, even under LL conditions, this work also provides insight into the role of enzyme transcript abundance in metabolic control. The interaction of carbon sources, reducing power, and enzyme concentration is important in the control of all metabolic networks; *S. elongatus* is an ideal organism for dissecting these interactions in a robust but non-steady state system.

RESULTS

We monitored metabolic concentrations via targeted liquid chromatography-tandem mass spectrometry (LC/MS/MS) in entrained cultures of *S. elongatus* (Kelly et al. 2011; Locasale et al. 2012; Yuan et al. 2012; Kelly et al. 2011; Locasale et al. 2012; Yuan et al. 2012). A full list of metabolites monitored is available in Table 4.1. For all experiments, *S. elongatus* was entrained by two consecutive 12hr:12hr LD periods (See Materials and Methods for details on entrainment and growth conditions). We first examined metabolic dynamics at a 2-hr resolution over a 24hr 12:12 LD period post entrainment. 107 of the 171 metabolites tracked over the 24hr period reached peak concentrations during the 12hr light period (Figure 4.1A, File S1A). This is consistent with metabolite abundances increasing in response to continual photosynthetic activity when light is present. K-means clustering (4 clusters, 1000 replicates) of these metabolites suggested that additional light-independent regulation also affected metabolite concentrations. Two of the four

clusters, representing over 45% of measured metabolites, reached maximal concentrations during discrete time windows (Figure 4.1B).

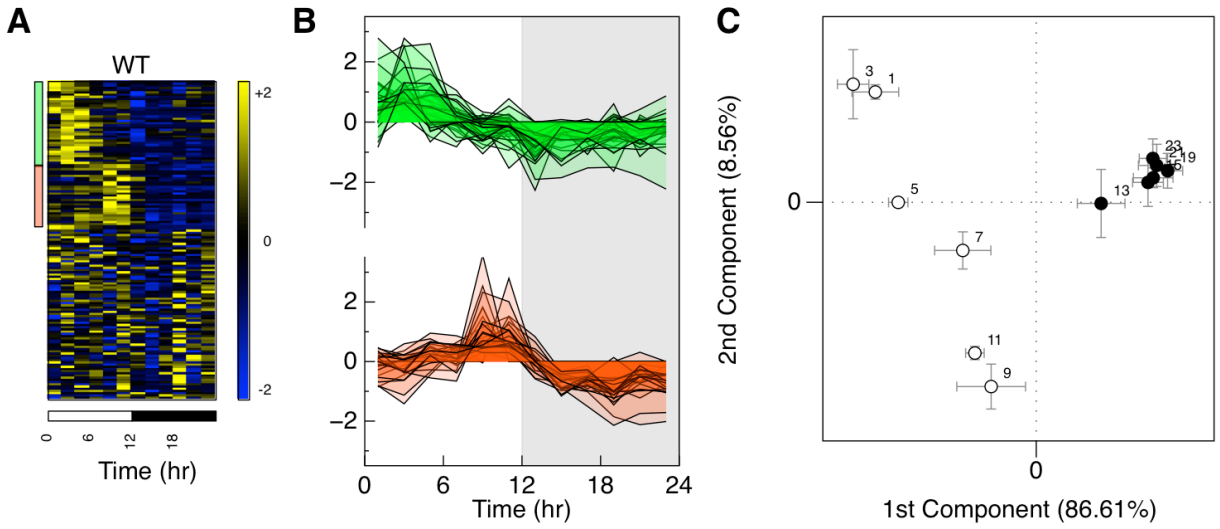


Figure 4.1 Metabolomic profiling of a 24-hr LD cycle. (A) Heatmap of 145 metabolites tracked (see Materials and Methods for details). K-means clustering revealed a cluster that peaks early in the day (green bar) and a cluster that peaks late in the day (orange bar). (B) Comparison of early (green) and late (orange) peaking clusters. See supplemental Table 4.2 for a listing of clustered metabolites. (C) PLSR analysis indicates that metabolic activity is regulated primarily while light is present. This is evidenced by the tight clustering of the samples collected in the dark (black circles) versus the time-dependent organization of the samples collected in the light (white circles). Axes are labeled with the percentage of the variance in the response variable explained by the PLSR model. Numbers indicate the circadian time of sample collection, error bars represent the standard deviation of the 3 technical replicates taken at each timepoint.

If there is light-dependent regulation of metabolite concentration in *S. elongatus*, metabolite concentrations in the 24 hr LD dataset should be predictive of lighting status. To test this assumption, we employed partial least squares regression analysis (PLSR). PLSR is a partially supervised technique that considers a multiparametric matrix of predictor values and uses that dataset to model a dependent matrix of response values. PLSR constructs orthogonal axes that best describe the variance of the predictor and response values (Wold et al. 2001). We applied two-component PLSR to the 24-hr dataset to predict lighting status as the response variable (see Materials and Methods for parameters). The resulting model (R^2 PLS = 95.16%) tightly clustered the

data collected during the dark, while spreading the light timepoints out roughly according to time of day in the first two principal components (Figure 4.1C).

The time-dependent spacing of the daytime timepoints in the PLSR model is a result of the underlying structure of the data, and not the lighting status variable, as no information on time of sample collection was incorporated into the model. The vast majority of the variance explained by PLSR is in the first component, which clusters early day and late day time-points separately. This indicates that *S. elongatus* metabolism is regulated primarily during light periods when the photosystems are active.

To distinguish between metabolites under circadian regulation versus those that simply respond to photosynthetic activity, we tracked metabolite dynamics for 48 hr at LL. In LL conditions, light-responsive metabolites are expected to remain fairly invariant in concentration, while circadian metabolites should oscillate in abundance with a 24-hr period. Under the conditions of the previous 24-hr experiment, cell division occurred approximately once every 12 hrs; we therefore increased the CO_2 concentration to 2% and the temperature to 35°C such that cell doubling time was reduced to approximately 8 hrs (see Materials and Methods). This was done to ensure that any 24-hr oscillations would be generated by circadian regulation rather than the cell cycle. To track the status of the Kai clock during these experiments, we integrated a bacterial luciferase operon under the control of the circadian *PsbAI* promoter into the genome (Kondo et al. 1993). For all subsequent experiments, wild-type (WT) S. elongatus includes this luciferase reporter integrated into the genome (see Materials and Methods for details). In addition, we also constructed strain Δ*KaiABC*, which differs from the WT reporter strain in that the entire *KaiABC* locus has been deleted (see Supporting Information for a full list of strains and plasmids).

During the 48-hr LL experiment, the luciferase reporter confirmed that the circadian clock maintained a 24-hr period in the WT cells, while the Δ*KaiABC* cells were arrhythmic (Figure 4.2A

and B). 217 metabolites were detected over the course of the experiment (File S1B). Unlike the 24hr LD dataset, unsupervised methods to cluster the WT metabolite data failed to identify obvious clusters of correlated metabolites (Figure 4.2C).

Three-component PLSR (R^2 PLS = 95.79) of the WT data over 48 hrs in LL provides insight into the role of light in metabolic concentrations (Figure 4.3A). For this analysis, the circadian luciferase signal was employed as the dependent response variable, such that metabolite data is used to predict Kai-dependent luciferase activity (see Materials and Methods for details). The first 2 PLSR components separate samples collected during the subjective day and night, while maintaining clustering of samples collected at the same circadian time. As a control, performing the same PLSR analysis on the Δ*KaiABC* data, again using the WT luciferase signal as the response variable, did not significantly separate the timepoints or identify subjective day/night transitions (Figure 4.5). An interesting outlier is the hr 0 sample, which was taken at the dark to light transition. This suggests that the activation of lighting at this timepoint had a greater effect on metabolite concentrations than the status of the circadian clock. Both the WT and Δ*KaiABC* samples demonstrated this behavior, further indicating that this effect is circadian-independent. Of the 216 metabolites measured in the WT sample, 107 reached their maximum at hr 0, as compared to 18 of 171 in the 24-hr LD experiment.

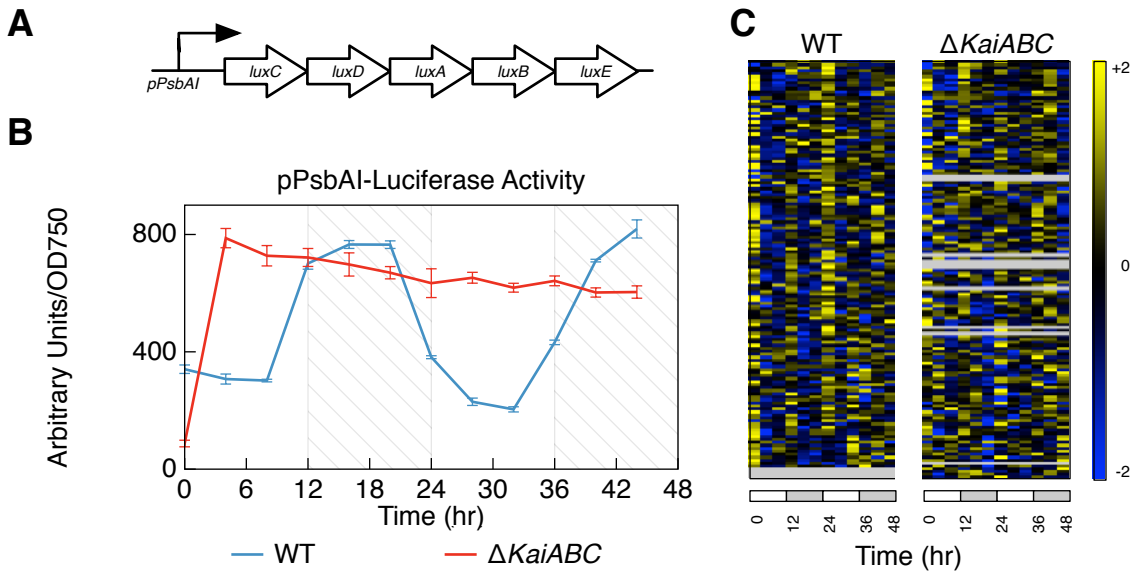


Figure 4.2 Metabolomic profiling of a 48-hr LL period. (A) Design of the circadian luciferase reporter used in this study. (B) Luciferase activity during the 48-hr measurement period. Two full oscillations of the circadian rhythm are evident in the WT sample. The luciferase activity in the Δ*KaiABC* sample responds to the dark to light transition at hr 0, then remains flat for the duration of the LL period. Error bars represent the standard deviation of 3 technical replicates taken at each timepoint. Diagonal lines represent subjective night hours. (C) Heatmap of 139 metabolites tracked. Metabolites are organized in descending order of the WT circadian score (see Materials and Methods).

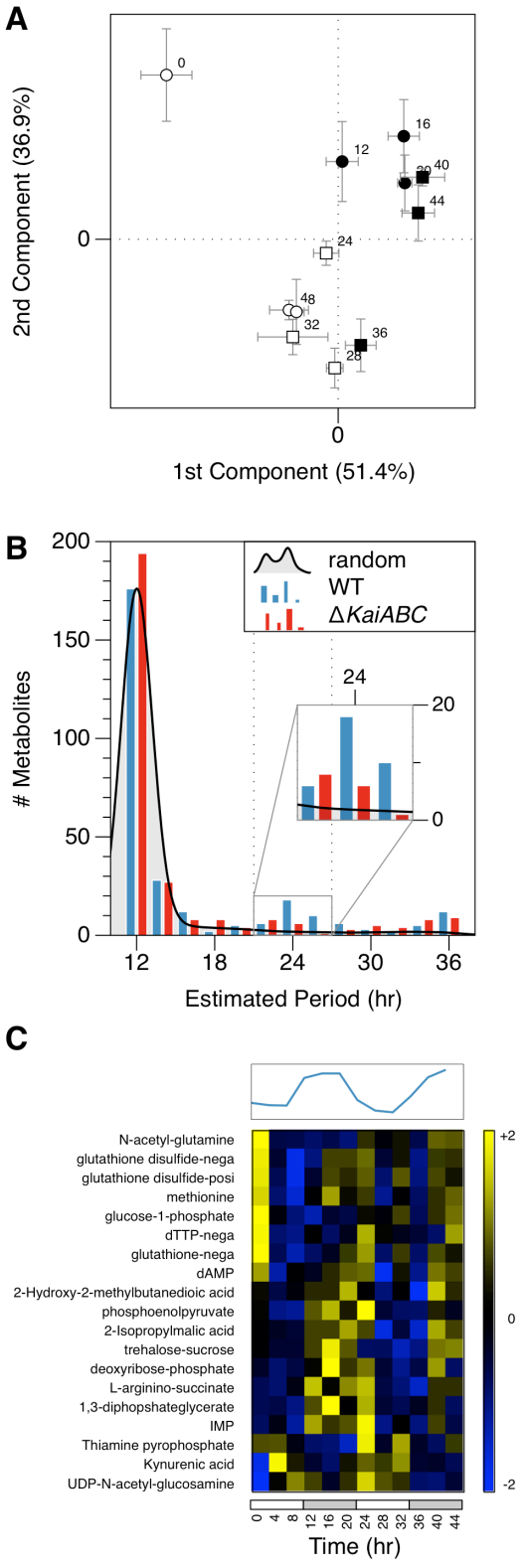


Figure 4.3 Circadian analysis of 48-hr LL data.

Figure 4.3, continued. Circadian analysis of 48-hr LL data. (A) PLSR analysis, using the WT luciferase signal as the response variable. Axes are labeled with the percentage of the variance in the response variable explained by the PLSR model. White and black markers indicate whether the sample was collected during subjective day or subjective night, respectively. Subjective day and subjective night samples are segregated, with samples collected near the subjective day/night transition near 0 on the 1st component axis. Circles and squares denote whether the sample was collected on day 1 or day 2 of the experiment, respectively. Timepoints and error bars are represented identically to Figure 4.1C. (B) Histogram of periods calculated for each metabolite via Cosiner analysis. Metabolites with periods less than or equal to 12 hrs or greater than 36 hours are considered arrhythmic by this analysis. The Δ*KaiABC* distribution closely matches that of an expected random distribution (see Materials and Methods for details). The WT distribution deviates from the random distribution, with a greater number of metabolites with periods between 22 and 26 hrs (inset). (C) WT metabolites that displayed circadian periodicity (see Materials and Methods for selection criteria). The WT luciferase signal from this experiment is reproduced above the heatmap.

To identify metabolites with apparent 24-hr periods, we applied a modified Cosiner algorithm to the 48-hr LL dataset (Vijayan et al. 2009; Kucho et al. 2005). In comparison to the Δ*KaiABC* sample, the WT sample possessed fewer metabolites with periods less than 12 hrs, the shortest period that we attempted to identify. In addition, there was a clear peak at 24 hrs in the WT histogram that was not present in the Δ*KaiABC* histogram (Figure 4.3B). The Δ*KaiABC* histogram resembles the expected period distribution of randomized samples, while the WT data clearly deviates from a random distribution (Figure 4.3B). We identified 19 putative circadian metabolites in the WT dataset via the Cosiner algorithm (Figure 4.3C). Running the same analysis on the Δ*KaiABC* data identified 4 putative circadian metabolites (File S1C); this result is likely to be representative of the false positive rate of circadian identification.

The most striking putative circadian trajectory belonged to sucrose/trehalose, which also peaked in the late day during the 24-hr LD experiment (Figure 4.4A). The metabolic precursor sucrose/trehalose-6-phosphate may also display circadian periodicity (Figure 4.4B), although this compound was not measured with sufficient fidelity to definitively support that observation. Currently, our instrumentation is unable to distinguish between sucrose and trehalose. Therefore, to determine which compound was detected, and to elucidate the potential role of that compound in

the circadian clock, we knocked out the enzymes responsible for their synthesis. Both sucrose and trehalose can be synthesized from UDP-glucose (Figure 4.4C). Sucrose phosphate synthase is encoded by *spsA* in cyanobacteria (Porchia & Salerno 1996; Curatti et al. 1998). Although no open reading frame in S. elongatus PCC 7942 is annotated as a trehalose phosphate synthase, the ORF *syc1355_d* in S. elongatus PCC 6301 has been annotated as a trehalose phosphate synthase. BLAST of *syc1355_d* against the S. elongatus PCC 7942 genome revealed that *syc1355_d* is nearly identical to the *dpm* gene in S. elongatus PCC 7942 (Altschul et al. 1990) (Figure 4.6). We therefore knocked out both *spsA* and *dpm* in the WT strain background with the PsbAI-luciferase reporter.

While the Δ*spsA* strain showed no significant disruption of *pPsbAI*-luciferase activity, the Δ*dpm* luciferase signal peaks earlier in the day upon release into LL conditions (Figures 4.4D and 4.7). Targeted LC/MS/MS of these strains and the WT strain revealed that the sucrose/trehalose concentration was decreased in the Δ*spsA* strain, and increased in the Δ*dpm* strain (Figure 4.4E, File S1D). This indicated that both sucrose and trehalose are present in S. elongatus. It also indicated that the circadian disruptions observed in the Δ*dpm* strain were likely to be the result of increased sucrose concentrations, not decreased trehalose concentrations, as evidenced by the three-fold increase in sucrose/trehalose observed in the Δ*dpm* strain. Furthermore, both gene deletions increased glucose-6-phosphate and fructose-6-phosphate concentrations, likely due to flux redistribution in the sucrose/trehalose synthesis pathways (Figure 4.4E). Remarkably, these metabolic differences were evident despite a relatively uniform *pPsbAI*-luciferase signal at the time of sample collection (Figure 4.4D, collection at t = 12.5 hr in LL). The metabolic impact of the *dpm* deletion disrupted circadian-dependent control of gene expression under LL conditions (Figure 4.4D). This result presents a potential role of circadian-regulated metabolites in reinforcing Kai clock entrainment in LL conditions.

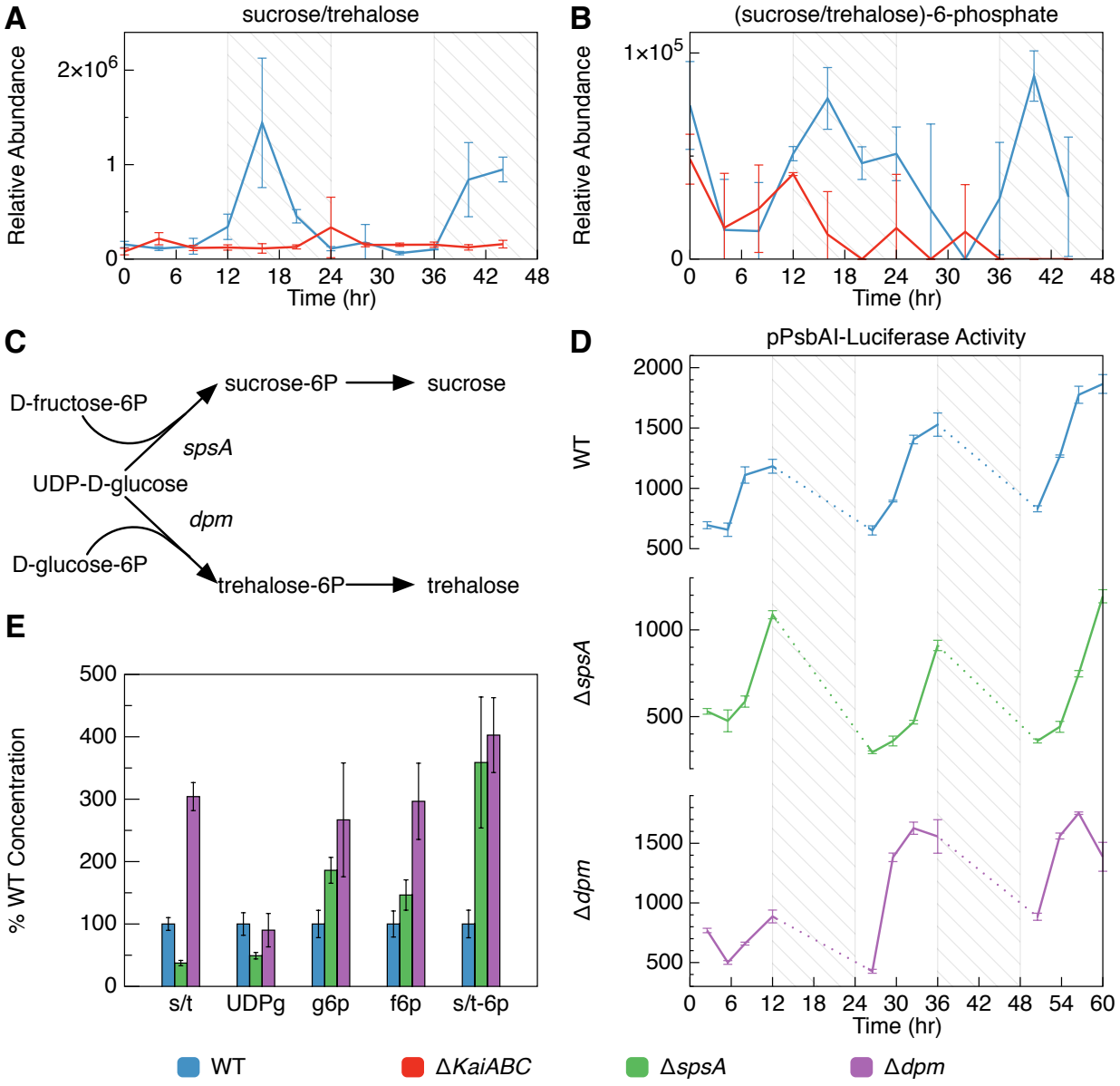


Figure 4.4 Disaccharide oscillations under LL conditions, and phenotypes of mutants defective in disaccharide synthesis. (A) Sucrose/trehalose and (B) sucrose-6-phosphate/trehalose-6-phosphate concentrations over the course of the 48-hour LL experiment. Units in both (A) and (B) are the calculated peak area for that SRM, normalized to the OD at 750 nm (See Materials and Methods). (C) Biosynthesis of sucrose and trehalose. In *S. elongatus*, *spsA* has been verified as the sole sucrose-phosphate synthase, while *dpm* is a putative trehalose-phosphate synthase. (D) Representative luciferase dynamics of Δ*spsA* and Δ*dpm* strains of *S. elongatus* with the *pPsbAI*-luciferase reporter. Time 0 represents the beginning of LL conditions following 2 LD entrainment periods. No measurements were made during subjective night hours, as indicated by dotted lines. Y-axes are arbitrary luciferase units normalized to the OD at 750 nm. (E) Selected sucrose and trehalose biosynthetic intermediates compared at 12.5 hrs in LL. Reported values are SRM peak area normalized to OD, expressed as a percentage of the area/OD in the WT sample. Metabolites were extracted from entrained cultures of WT, Δ*spsA,* and Δ*dpm,* and analyzed

Figure 4.4, continued. via targeted LC/MS/MS. Full metabolite data from this experiment are included as File S1D. The luciferase data from this experiment is presented in Figure 4.7. s/t = sucrose/trehalose, UDPg = UDP-D-glucose, g6p = glucose-6-phosphate, f6p = fructose-6-phosphate, s/t-6p = sucrose/trehalose-6-phosphate.

DISCUSSION

By observing metabolic dynamics in entrained cultures of *S. elongatus* under LD and LL conditions, we isolated the influence of the circadian rhythm from that of photosynthetic activity on the metabolic network. A high proportion of circadian metabolites under LL conditions would have implied that circadian oscillations in gene expression directly regulate the metabolic network. Instead, we observed that while time dependent regulation is evident under LD conditions, metabolite fluctuations are largely non-circadian under LL conditions. These data demonstrated that much of S. elongatus metabolism is "upstream" of the circadian clock and relatively insensitive to transcriptional regulation. This finding has important implications for the metabolic engineering of cyanobacteria, since many metabolites do not vary as a function of the global oscillations in gene expression. Interestingly, *spsA* and *dpm* transcript levels under LL conditions do resemble the observed sucrose/trehalose oscillations (Ito et al. 2009; Vijayan et al. 2009)(Figure 4.8). This indicates that enzyme expression oscillations can contribute to metabolite oscillations, but that expression oscillations alone are insufficient to cause this behavior.

The S. elongatus circadian system is ideal for investigating the consequences of transcriptional variation on intracellular metabolites. Earlier work to unite transcriptome and metabolome data from Escherichia coli (*E. coli*) demonstrated that perturbation of enzyme expression does not necessarily impact the concentrations of associated metabolites. Instead, it was observed that enzyme deletions often lead to re-routing of metabolic flux without appreciable changes in enzyme or metabolite concentrations (Ishii et al. 2007). In S. elongatus, gene expression does not reach a single steady-state since most gene transcripts are under circadian control (Ito et al.

2009; Vijayan et al. 2009). Likewise, metabolite levels also do not reach a steady-state under LD conditions. The combination of circadian and light control results in the discrete regulation of metabolite concentrations, as evidenced in our 24-hr LD experiment (Figure 4.1). The opportunity to abolish circadian regulation of transcription then allowed us to determine to what extent transcriptional regulation influenced metabolite levels.

Our data indicate that circadian-regulated metabolites such as sucrose may be serving to reinforce the circadian rhythm when the lighting rhythm is disrupted (Figure 4.4D). However, these metabolites do not appear to entrain the Kai oscillator itself. This agrees with previous studies demonstrating that intracellular redox status, which is largely determined by the presence of light and associated photosynthesis, entrains the circadian clock in S. elongatus (Rust et al. 2011). Sucrose may improve the robustness of Kai oscillations by stabilizing the energy state (i.e. ATP/ADP and NADH/NAD+ ratios) oscillations that occur under a diurnal lighting rhythm. Alternatively, disaccharides such as sucrose and trehalose have been identified as osmoprotectants in S. elongatus (Blumwald et al. 1983; Niederholtmeyer et al. 2010). Modulation of chromosomal superhelicity has been shown to be part of the osmotic stress response in E. coli (Cheung et al. 2003). In S. elongatus, chromosomal superhelicity is under circadian control (Vijayan et al. 2011); sucrose oscillations may reinforce the periodicity of superhelicity in LL conditions. Interestingly, sucrose has been identified as a circadian regulator in higher plants (Dalchau et al. 2011). Cyanobacteria and plants both conduct oxygenic photosynthesis but differ greatly in their circadian clock architectures (Mackey 2007). Sucrose oscillations may be a common feature of circadian-regulated photosynthetic metabolism.

With the exception of the candidate metabolites identified in this study, the contribution of circadian oscillations in gene expression does not appear to result in circadian metabolic oscillations. This is surprising, given that as many as 70% of metabolic enzymes appear to be transcriptionally controlled by the circadian clock in S. elongatus (Vijayan et al. 2009). This has implications for the

modeling of metabolic control; transcriptome data is insufficient to understand metabolic dynamics, requiring direct measurement of the metabolome.

Cyanobacteria such as *S. elongatus* show great promise as a platform for solar-driven metabolic engineering (Ducat et al. 2011). To better realize this goal, our data can be employed to time the expression of exogenous enzymes with the availability of endogenous substrates. Together with work to analyze the spatial dynamics of *S. elongatus* metabolism (Savage et al. 2010), our results contribute towards a spatiotemporal atlas of photoautotrophic metabolism.

MATERIALS AND METHODS

Strain Construction

All relevant strains, plasmids, and oligos are described in Tables S3-5. Wild-type *Synechococcus elongatus* PCC 7942 was acquired from the American Type Culture Collection (ATCC). *S. elongatus* cells were transformed following standard protocols by incubating 10^8 cells overnight in the dark with 100 ng of plasmid DNA and immediately plating on selective media (Clerico et al. 2007).

The *psbAI* promoter was amplified from *S. elongatus* genomic DNA, and cloned in front of the *luxCDABE* operon from *Photorhabdus luminescens* (Kishony & Leibler 2003). This luciferase construct was cloned into pAM2314 with EcoRI and NotI. This vector integrates gene constructs into Neutral Site I in the *S. elongatus* genome, a locus that has been shown to permit gene integration without affecting strain phenotype (Mackey et al. 2007; Savage et al. 2010).

Gene knockouts were generated by in-frame deletion of the intended gene using an antibiotic resistance cassette. Gene knockout plasmids were constructed as follows. 800 bp flanking regions for each open reading frame to be deleted were amplified from *S. elongatus* genomic DNA using a BioBrick-like strategy in which BioBrick restriction sites were added to each part (Phillips & P. Silver 2006). A table of DNA oligonucleotides used in cloning is shown in Table 4.5. Briefly,

EcoRI and SpeI were used for upstream restriction sites and HindIII and NotI were used as downstream sites. The final part was therefore EcoRI – upstream flanking sequence SpeI – antibiotic resistance cassette – HindIII – downstream flanking sequence – NotI. This part was ligated into the V0120 vector digested with EcoRI and NotI. For the deletion of the KaiABC locus, the hygromycin resistance cassette cloned from pUV15tetORm (Addgene) was used. For all other deletions, a chloramphenicol resistance cassette cloned from plasmid pACYCDuet-1 (Novagen) was used. Knockouts were created by transforming *S. elongatus* using the appropriate plasmid as described above and verified by colony PCR on single colonies isolated following colony purification of transformed strains.

Growth Conditions

For the 24-hr LD experiment, *S. elongatus* was grown in BG11 media at pH 8.8 (SIGMA) at 25°C under illumination of 3000 lux (~80 μE m^2 s^{-1}) via F15T8 fluorescent bulbs (Gro-Lux, 15W; Sylvania). Antibiotic concentrations were 10 μg/ml for each antibiotic. Cultures were stirred at 200 rpm via a magnetic stir bar. For all subsequent experiments, *S. elongatus* was grown at 35°C with CO_2 controlled at 2% in a Multitron shaking incubator (Infors HT). Shaking was set to 150 rpm, and illumination parameters remained at 3000 lux (~80 μE m^2 s^{-1}) via F15T8 fluorescent bulbs (Gro-Lux, 15W; Sylvania). BG11 media (SIGMA) was supplemented with 1 g/L HEPES and adjusted to pH 8.8 to prevent media acidification by the increased CO_2 concentration.

Sample Collection for LC/MS/MS Analysis

Cells were harvested at an OD750 of 0.5-0.8. For the 24-hr LD experiment, 1.8 ml of cells were collected, and centrifuged immediately at 21,000 rcf for 1 min in a microcentrifuge. The resulting pellets were then resuspended in 500 uL 80% (v/v) methanol at dry ice temperatures. Following brief vortexing, the samples were incubated on dry ice for 10 min. The samples were then centrifuged at 21,000 rcf for 1 min in a microcentrifuge at 4°C. The resulting supernatant was saved,

and the pellets were re-suspended in 250 uL 80% (v/v) methanol at dry ice temperatures. Vortexing, dry ice incubation, and centrifugation were repeated twice more; the resulting supernatants for each sample were pooled to a 1 mL final volume. For all subsequent experiments, 4 mL of cells were transferred to 16 mL 100% methanol, and centrifuged immediately at 4,000 rcf for 5 min at 4°C. The resulting pellets were then resuspended in 500 uL 80% (v/v) methanol at dry ice temperatures, and the remaining extraction protocol proceeded as in the 24-hr LD experiment. All pooled supernatants were stored at -80°C immediately following collection.

Targeted LC/MS/MS Analysis

Samples were evaporated to dryness in a refrigerated speed vac, and resuspended using 20 µl high performance liquid chromatography (HPLC)-grade water for mass spectrometry. We injected 10 µl and analyzed it using a 5500 QTRAP triple quadrupole mass spectrometer (AB/SCIEX) coupled to a Prominence HPLC system (Shimadzu) using selected reaction monitoring (SRM) of a total of 249 endogenous water soluble metabolites for analyses of samples. Some metabolites were targeted in both the positive and negative ion mode for a total of 297 SRM transitions using positive/negative ion polarity switching. ESI voltage was 5,000 V in the positive ion mode and −4,500 V in the negative ion mode. The dwell times were 4 ms per SRM transition, and the total cycle time was 1.89 s. Samples were delivered to the MS using a 2.0 mm internal diameter × 15 cm Luna NH2 hydrophilic interaction chromatography (HILIC) column (Phenomenex) at 300 µl/min. Gradients were run starting from 85% buffer B (HPLC-grade acetonitrile) to 42% buffer B from 0–5 min; 42% buffer B to 0% buffer B from 5–16 min; 0% buffer B held from 16–24 min; 0% buffer B to 85% buffer B from 24–25 min; and 85% B held for 7 min to re-equilibrate the column. Buffer A was comprised of 20 mM ammonium hydroxide and 20 mM ammonium acetate in 95:5 water:acetonitrile. Peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant v1.1 software (Applied Biosystems).

Metabolites with defined SRM are listed in Table 4.1. Metabolite peak area data for all experiments is reported in File S1. The number of metabolites reported as detected in each experiment is the number of metabolites with a quantifiable peak area in at least one timepoint. For heatmap visualization and Cosiner analysis, we required that a metabolite be detected in all timepoints.

Luciferase Measurement

Triplicate samples of each culture were transferred to a white, clear bottom 96-well plate (Corning, product #3632) and incubated in the dark for 10 minutes at room temperature. After brief shaking in the dark, optical density at 750 nm and luciferase activity (3 s observation) were measured on a SpectraMax M5 Microplate Reader (Molecular Devices).

Circadian Period Estimation

Raw integrated peak data from the LC/MS/MS analysis for each sample in the 48-hr LL experiment was imported into MATLAB (Mathworks) and normalized by the OD 750 at each time point. Log2 transformed and linearly detrended data was analyzed via the Cosiner method (Vijayan et al. 2009; Kucho et al. 2005). 241 cosine waves, with periods between 12 and 36 hrs at 0.1-hr increments, were used for the comparison. For each metabolite trace, the period was assigned based on the cosine wave with the minimum Euclidean distance from that trace (Kucho et al. 2005). As in reference (Kucho et al. 2005), we required that each putative circadian metabolite have a Cosiner error factor ($Ef < 0.2$). To increase the stringency of this method, we made the following additional requirements: that the estimated period be between 22 and 26 h (versus 18 to 26.8 h in reference Kucho et al. 2005). We also required that the peak to trough distance for each day (over the 2 day experiment) was different to a significance of $p < 0.1$. We also generated 10^5 randomly permuted metabolite traces from the 48-hr data for comparison (Ito et al. 2009). We required putative circadian metabolites to have periods closer to 24 hrs than 99% of the metabolites in the random

set, i.e. an empirical $p < 0.1$. 19 metabolites in the WT sample and 4 metabolites in the Δ*KaiABC* sample met these criteria. A circadian score was assigned to each metabolite, equal to the average peak to trough p value for each day multiplied by the p value of the metabolite versus the random set. This score was used to determine the sorting order of WT metabolites in Figure 4.2C. All values calculated in these analyses are included in File S1C.

Partial Least Squares Regression (PLSR)

PLSR analysis was performed as described in (Wold et al. 2001) via the MATLAB plsregress function. The predictor matrix X contained the log2 transformed metabolite data for the given experiment. The response matrix Y contained the lighting status (0 = dark, 1 = light) for the 24-hr LD experiment, and the WT luciferase status (OD-adjusted luciferase units as shown in Figure 4.2B) for the 48-hr LL experiment. The number of components selected (2 components for the 24-hr LD experiment and 3 components for the 48-hr LL experiment) was based on minimization of mean squared error and 2-way cross-validation.

ACKNOWLEDGEMENTS

We thank Susan Golden for providing the pAM2314 plasmid. We also thank Devin Burrill, Mara Inniss, Danny Ducat, and Jeff Way for critical review of the manuscript. P.M.B. was supported by fellowships from the Harvard University Center for the Environment and the NSF Synthetic Biology Engineering Research Center. D.F.S. was a U.S. Department of Energy Biosciences fellow of the Life Sciences Research Foundation. P.A.S. acknowledges support from the Radcliffe Institute of Advanced Study, the Wyss Institute for Biologically Inspired Engineering, and the DOD Department of the Army.

SUPPLEMENTAL FIGURES AND TABLES

File S1 is available upon request and includes the following data. This dataset will be included as online supporting information when this work is published.

File S1A. Full LC/MS/MS data from 24-hr LD experiment
File S1B. Full LC/MS/MS data from 48-hr LL experiment
File S1C. Results of Cosiner analysis from 48-hr LL experiment
File S1D. Full LC/MS/MS data from WT, Δ*spsA*, Δ*dpm* experiment

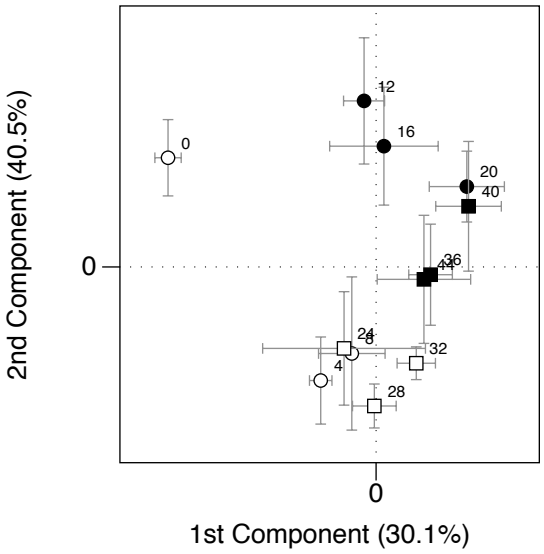


Figure 4.5 PLSR of 48-hr LL Δ*KaiABC* data with WT luciferase data as the response variable. While this PLSR model superficially separates subjective day and night samples, some separation is expected due to the use of the luciferase signal from the WT data as the response variable. Since the separation between subjective day and night samples is not significant, this indicates that the Δ*KaiABC* data is a poor predictor of the WT luciferase status.

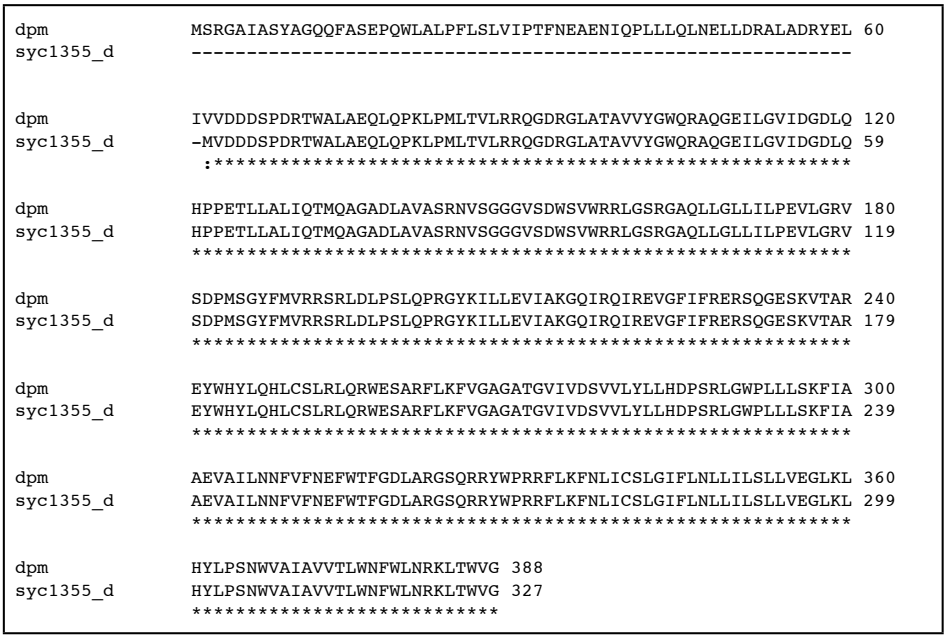


Figure 4.6 Sequence alignment of *S. elongatus* PCC 6301 syc1355_d vs. *S. elongatus* PCC 7942 dpm. Aligment performed via the ClustalW2 multiple sequence alignment tool (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

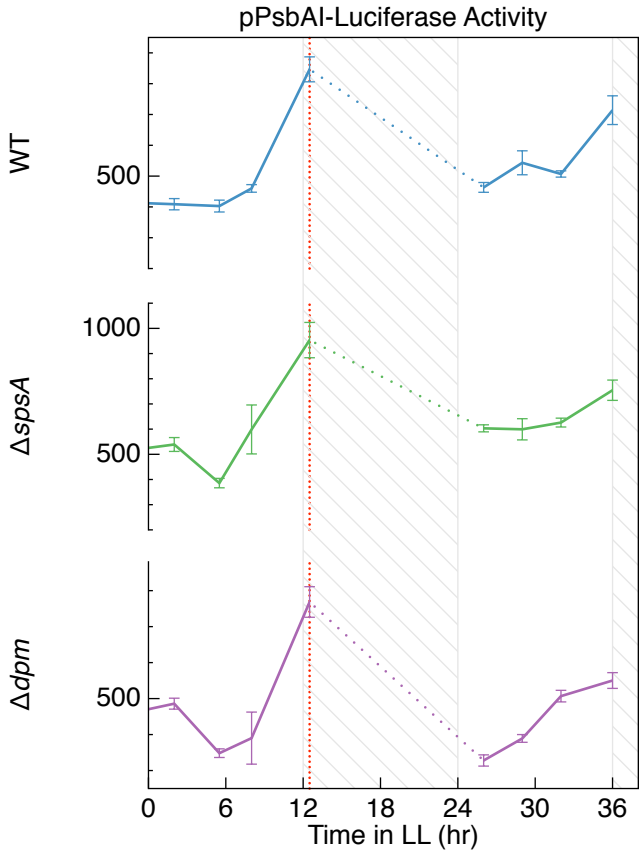


Figure 4.7 p*PsbAI*-luciferase data corresponding to metabolites measured in Figure 4.4E. Metabolite samples graphed in Figure 4.4E were collected 12.5 hrs into the LL period, as indicated by the dotted red line. No measurements were made during subjective night hours, as indicated by dotted lines. Y-axes are arbitrary luciferase units normalized to the OD at 750 nm. As in Figure 4.4D, the Δ*dpm* luciferase signal begins to peak earlier in the subjective day after 24 hours in LL.

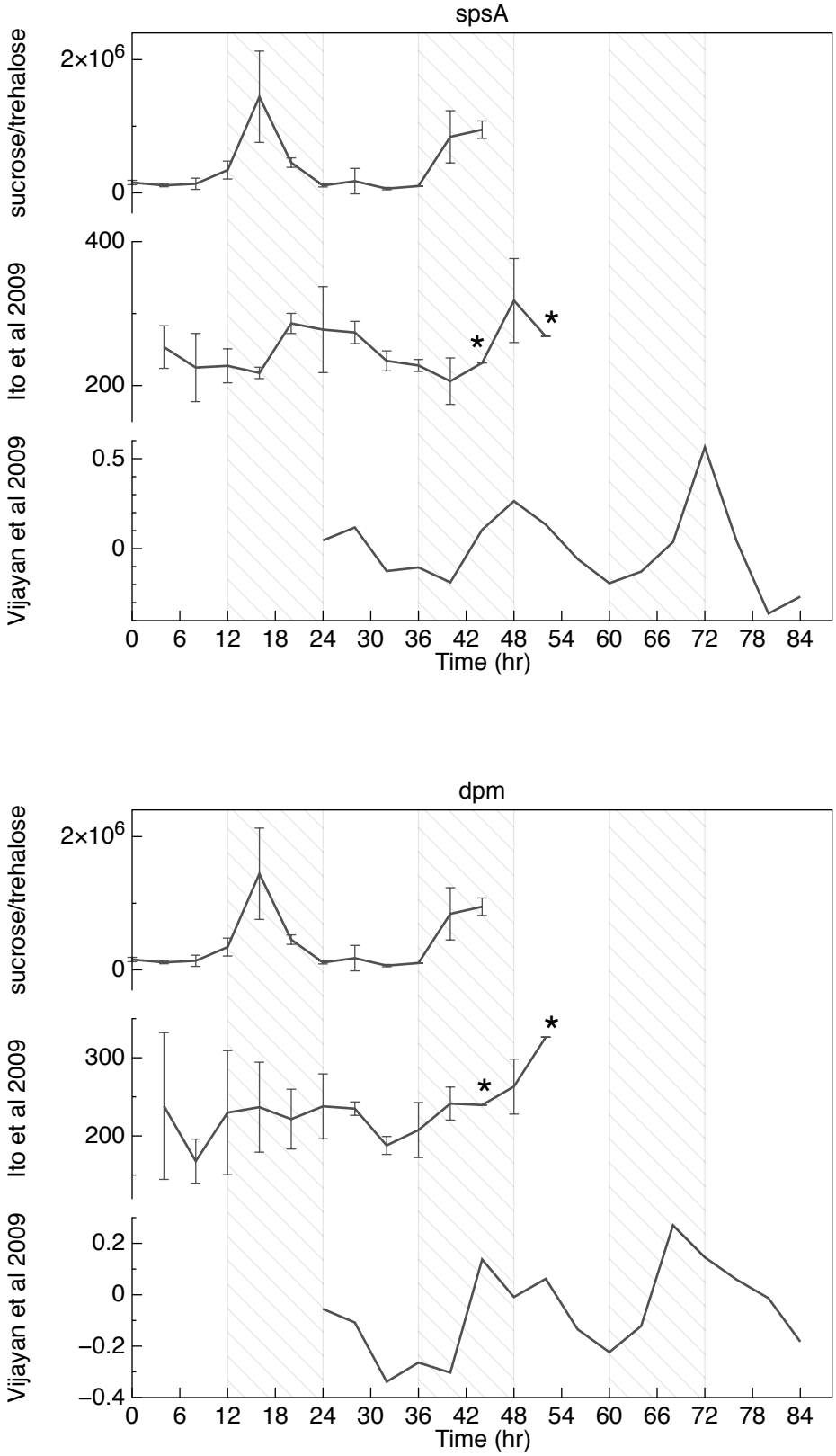


Figure 4.8 Comparison of enzyme expression levels from Ito et al 2009 and Vijayan et al 2009 to selected metabolites.

Figure 4.8, continued. Comparison of enzyme expression levels from Ito et al 2009 and Vijayan et al 2009 to selected metabolites. Asterisks denote time points in the Ito et al data where only a single sample was collected. For all other time points, n=2 for Ito et al data, and n=1 for Vijayan et al data.

Table 4.1. List of metabolites monitored in each experiment

24-hr LD experiment	48-hr LL experiment	WT, ΔspsA, Δdpm hr 12.5
glyoxylate	glyoxylate	glyoxylate
glycolate	glycolate	glycolate
pyruvate	pyruvate	pyruvate
lactate	lactate	lactate
2-oxobutanoate	2-oxobutanoate	2-oxobutanoate
acetoacetate	acetoacetate	acetoacetate
glycerate	glycerate	glycerate
uracil	uracil	uracil
fumarate	fumarate	fumarate
Maleic acid	Maleic acid	Maleic acid
2.keto-isovalerate	2-keto-isovalerate	2-keto-isovalerate
Guanidoacetic acid	Guanidoacetic acid	Guanidoacetic acid
succinate	succinate	succinate
Methylmalonic acid	Methylmalonic acid	Methylmalonic acid
3.S-methylthiopropionate	3-S-methylthiopropionate	3-S-methylthiopropionate
nicotinate	nicotinate	nicotinate
taurine	taurine	taurine
Pyroglutamic acid	Pyroglutamic acid	Pyroglutamic acid
Citraconic acid	Citraconic acid	Citraconic acid
2-ketohaxanoic acid	2-ketohaxanoic acid	2-ketohaxanoic acid
N.Acetyl.L-alanine	N-Acetyl-L-alanine	N-Acetyl-L-alanine
oxaloacetate	oxaloacetate	oxaloacetate
Hydroxyisocaproic acid	Hydroxyisocaproic acid	Hydroxyisocaproic acid
malate	malate	malate
hypoxanthine	hypoxanthine	hypoxanthine
anthranilate	anthranilate	anthranilate
p-aminobenzoate	p-aminobenzoate	p-aminobenzoate
p-hydroxybenzoate	p-hydroxybenzoate	p-hydroxybenzoate
acetylphosphate	acetylphosphate	acetylphosphate
Carbamoyl phosphate	Carbamoyl phosphate	Carbamoyl phosphate
a-ketoglutarate	a-ketoglutarate	a-ketoglutarate
Phenylpropiolic acid	Phenylpropiolic acid	Phenylpropiolic acid
2.oxo.4-methylthiobutanoate	2-oxo-4-methylthiobutanoate	2-oxo-4-methylthiobutanoate
2.Hydroxy.2-methylbutanedioic acid	2-Hydroxy-2-methylbutanedioic acid	2-Hydroxy-2-methylbutanedioic acid
3-methylphenylacetic acid	3-methylphenylacetic acid	3-methylphenylacetic acid
xanthine	xanthine	xanthine
Hydroxyphenylacetic acid	Hydroxyphenylacetic acid	Hydroxyphenylacetic acid
2,3-dihydroxybenzoic acid	2,3-dihydroxybenzoic acid	2,3-dihydroxybenzoic acid
orotate	orotate	orotate
dihydroorotate	dihydroorotate	dihydroorotate
allantoin	allantoin	allantoin

Table 4.1, continued

Aminoadipic acid	Aminoadipic acid	Aminoadipic acid
Indole.3-carboxylic acid	Indole-3-carboxylic acid	Indole-3-carboxylic acid
phenylpyruvate	phenylpyruvate	phenylpyruvate
Atrolactic acid	Atrolactic acid	Atrolactic acid
Phenyllactic acid	Phenyllactic acid	Phenyllactic acid
quinolinate	quinolinate	quinolinate
phosphoenolpyruvate	phosphoenolpyruvate	phosphoenolpyruvate
Uric acid	Uric acid	Uric acid
dihydroxy.acetone-phosphate	dihydroxy-acetone-phosphate	dihydroxy-acetone-phosphate
D.glyceraldehdye.3-phosphate	D-glyceraldehdye-3-phosphate	D-glyceraldehdye-3-phosphate
sn.glycerol.3-phosphate	sn-glycerol-3-phosphate	sn-glycerol-3-phosphate
shikimate	shikimate	shikimate
aconitate	aconitate	aconitate
allantoate	allantoate	allantoate
Ascorbic acid	Ascorbic acid	Ascorbic acid
2-Isopropylmalic acid	2-Isopropylmalic acid	2-Isopropylmalic acid
N.carbamoyl.L.aspartate-nega	N-carbamoyl-L-aspartate-nega	N-carbamoyl-L-aspartate-nega
Pyrophosphate	Pyrophosphate	Pyrophosphate
glucono.?-lactone	glucono-?-lactone	glucono-?-lactone
myo-inositol	myo-inositol	myo-inositol
hydroxyphenylpyruvate	hydroxyphenylpyruvate	hydroxyphenylpyruvate
homocysteic acid	homocysteic acid	homocysteic acid
4-Pyridoxic acid	4-Pyridoxic acid	4-Pyridoxic acid
3-phosphoglycerate	3-phosphoglycerate	3-phosphoglycerate
Indoleacrylic acid	Indoleacrylic acid	Indoleacrylic acid
Kynurenic acid	Kynurenic acid	Kynurenic acid
citrate-isocitrate	citrate-isocitrate	citrate-isocitrate
isocitrate	isocitrate	isocitrate
citrate	citrate	citrate
2.dehydro.D-gluconate	2-dehydro-D-gluconate	2-dehydro-D-gluconate
D-gluconate	D-gluconate	D-gluconate
D.erythrose.4-phosphate	D-erythrose-4-phosphate	D-erythrose-4-phosphate
Xanthurenic acid	Xanthurenic acid	Xanthurenic acid
lipoate	lipoate	lipoate
D-glucarate	D-glucarate	D-glucarate
deoxyribose-phosphate	deoxyribose-phosphate	deoxyribose-phosphate
pantothenate	pantothenate	pantothenate
prephenate	prephenate	prephenate
deoxyuridine	deoxyuridine	deoxyuridine
ribose-phosphate	ribose-phosphate	ribose-phosphate
thymidine	thymidine	thymidine
uridine	uridine	uridine
deoxyinosine	deoxyinosine	deoxyinosine

Table 4.1, continued

shikimate.3-phosphate	shikimate-3-phosphate	shikimate-3-phosphate
D.glucono.?.lactone.6-phosphate	D-glucono-?-lactone-6-phosphate	D-glucono-?-lactone-6-phosphate
hexose-phosphate	hexose-phosphate	hexose-phosphate
glucose.1-phosphate	glucose-1-phosphate	glucose-1-phosphate
glucose.6-phosphate	glucose-6-phosphate	glucose-6-phosphate
fructose.6-phosphate	fructose-6-phosphate	fructose-6-phosphate
1,3-diphopshateglycerate	1,3-diphopshateglycerate	1,3-diphopshateglycerate
2,3-Diphosphoglyceric acid	2,3-Diphosphoglyceric acid	2,3-Diphosphoglyceric acid
S.ribosyl.L.homocysteine-nega	S-ribosyl-L-homocysteine-nega	S-ribosyl-L-homocysteine-nega
inosine	inosine	inosine
6.phospho.D-gluconate	6-phospho-D-gluconate	6-phospho-D-gluconate
xanthosine	xanthosine	xanthosine
D.sedoheptulose.1.7-phosphate	D-sedoheptulose-1-7-phosphate	D-sedoheptulose-1-7-phosphate
N.acetyl.glucosamine.1-phosphate	N-acetyl-glucosamine-1-phosphate	N-acetyl-glucosamine-1-phosphate
glutathione-nega	glutathione-nega	glutathione-nega
dUMP-nega	dUMP-nega	dUMP-nega
Geranyl-PP	Geranyl-PP	Geranyl-PP
O8P-O1P	O8P-O1P	O8P-O1P
dTMP-nega	dTMP-nega	dTMP-nega
cyclic-AMP	cyclic-AMP	cyclic-AMP
fructose.1,6-bisphosphate	fructose-1,6-bisphosphate	fructose-1,6-bisphosphate
trehalose-sucrose	trehalose-sucrose	trehalose-sucrose
Cellobiose	Cellobiose	Cellobiose
orotidine.5-phosphate	orotidine-5-phosphate	orotidine-5-phosphate
SBP	SBP	SBP
trans, trans-farnesyl diphosphate	trans, trans-farnesyl diphosphate	trans, trans-farnesyl diphosphate
S.adenosyl.L.homocysteine-nega	S-adenosyl-L-homocysteine-nega	S-adenosyl-L-homocysteine-nega
dCDP-nega	dCDP-nega	dCDP-nega
5.phosphoribosyl.1-pyrophosphate	5-phosphoribosyl-1-pyrophosphate	5-phosphoribosyl-1-pyrophosphate
Deoxycholic acid	Deoxycholic acid	Deoxycholic acid
OBP	OBP	OBP
dTDP-nega	dTDP-nega	dTDP-nega
CDP-nega	CDP-nega	CDP-nega
UDP-nega	UDP-nega	UDP-nega
Cholic acid	Cholic acid	Cholic acid
trehalose.6-Phosphate	trehalose-6-Phosphate	trehalose-6-Phosphate
Thiamine pyrophosphate	Thiamine pyrophosphate	Thiamine pyrophosphate
adenosine 5-phosphosulfate	adenosine 5-phosphosulfate	adenosine 5-phosphosulfate
ADP-nega	ADP-nega	ADP-nega
dGDP-nega	dGDP-nega	dGDP-nega
IDP-nega	IDP-nega	IDP-nega
GDP-nega	GDP-nega	GDP-nega
CDP-ethanolamine	CDP-ethanolamine	CDP-ethanolamine

Table 4.1, continued

FMN	FMN	FMN
cholesteryl sulfate	cholesteryl sulfate	cholesteryl sulfate
dCTP-nega	dCTP-nega	dCTP-nega
dUTP-nega	dUTP-nega	dUTP-nega
dTTP-nega	dTTP-nega	dTTP-nega
CTP-nega	CTP-nega	CTP-nega
UTP-nega	UTP-nega	UTP-nega
CDP-choline	CDP-choline	CDP-choline
dATP-nega	dATP-nega	dATP-nega
Taurodeoxycholic acid	Taurodeoxycholic acid	Taurodeoxycholic acid
ATP-nega	ATP-nega	ATP-nega
dGTP	dGTP	dGTP
GTP-nega	GTP-nega	GTP-nega
UDP.D-glucose	UDP-D-glucose	UDP-D-glucose
UDP.D-glucuronate	UDP-D-glucuronate	UDP-D-glucuronate
ADP.D-glucose	ADP-D-glucose	ADP-D-glucose
guanosine 5.diphosphate,3-diphosphate	guanosine 5-diphosphate,3-diphosphate	guanosine 5-diphosphate,3-diphosphate
UDP.N.acetyl-glucosamine	UDP-N-acetyl-glucosamine	UDP-N-acetyl-glucosamine
glutathione disulfide-nega	glutathione disulfide-nega	glutathione disulfide-nega
NAD+_nega	NAD+_nega	NAD+_nega
NADH-nega	NADH-nega	NADH-nega
dephospho.CoA-nega	dephospho-CoA-nega	dephospho-CoA-nega
cyclic bis(3->5) dimeric GMP	cyclic bis(3->5) dimeric GMP	cyclic bis(3->5) dimeric GMP
NADP+_nega	NADP+_nega	NADP+_nega
NADPH-nega	NADPH-nega	NADPH-nega
coenzyme A-nega	coenzyme A-nega	coenzyme A-nega
acetyl.CoA-nega	acetyl-CoA-nega	acetyl-CoA-nega
propionyl.CoA-nega	propionyl-CoA-nega	propionyl-CoA-nega
butyryl-CoA	butyryl-CoA	butyryl-CoA
acetoacetyl.CoA-nega	acetoacetyl-CoA-nega	acetoacetyl-CoA-nega
malonyl.CoA-nega	malonyl-CoA-nega	malonyl-CoA-nega
3.hydroxybutyryl-CoA	3-hydroxybutyryl-CoA	3-hydroxybutyryl-CoA
succinyl.CoA.methylmalonyl.CoA-nega	succinyl-CoA-methylmalonyl-CoA-nega	succinyl-CoA-methylmalonyl-CoA-nega
3.hydroxy.3.methylglutaryl.CoA-nega	3-hydroxy-3-methylglutaryl-CoA-nega	3-hydroxy-3-methylglutaryl-CoA-nega
Urea	2-hydroxygluterate	2-hydroxygluterate
ethanolamine	Urea	2-deoxyglucose-6-phosphate
Imidazole	ethanolamine	Urea
glycine	Imidazole	ethanolamine
alanine	glycine	Imidazole
betaine aldehyde	alanine	glycine
choline	betaine aldehyde	alanine
4-aminobutyrate	choline	betaine aldehyde
dimethylglycine	4-aminobutyrate	choline

Table 4.1, continued

serine	dimethylglycine	4-aminobutyrate
cytosine	serine	dimethylglycine
Creatinine	cytosine	serine
proline	Creatinine	cytosine
indole	proline	Creatinine
betaine	indole	proline
valine	betaine	indole
threonine	valine	betaine
homoserine	threonine	valine
purine	homoserine	threonine
cysteine	purine	homoserine
creatine	cysteine	purine
nicotinamide	creatine	cysteine
Imidazoleacetic acid	nicotinamide	creatine
thymine	Imidazoleacetic acid	nicotinamide
DL-Pipecolic acid	thymine	Imidazoleacetic acid
N-Acetylputrescine	DL-Pipecolic acid	thymine
hydroxyproline	N-Acetylputrescine	DL-Pipecolic acid
leucine-isoleucine	hydroxyproline	N-Acetylputrescine
ornithine	leucine-isoleucine	hydroxyproline
asparagine	ornithine	leucine-isoleucine
aspartate	asparagine	ornithine
adenine	aspartate	asparagine
Methylcysteine	adenine	aspartate
homocysteine	Methylcysteine	adenine
methylnicotinamide	homocysteine	Methylcysteine
histidinol	methylnicotinamide	homocysteine
lysine	histidinol	methylnicotinamide
glutamine	lysine	histidinol
O-acetyl-L-serine	glutamine	lysine
glutamate	O-acetyl-L-serine	glutamine
methionine	glutamate	O-acetyl-L-serine
guanine	methionine	glutamate
histidine	guanine	methionine
2-Aminooctanoic acid	histidine	guanine
carnitine	2-Aminooctanoic acid	histidine
Methionine sulfoxide	carnitine	2-Aminooctanoic acid
phenylalanine	Methionine sulfoxide	carnitine
Pyridoxamine	phenylalanine	Methionine sulfoxide
pyridoxine	Pyridoxamine	phenylalanine
1-Methyl-Histidine	pyridoxine	Pyridoxamine
N-acetyl-L-ornithine	1-Methyl-Histidine	pyridoxine
arginine	N-acetyl-L-ornithine	1-Methyl-Histidine

Table 4.1, continued

citrulline	arginine	N-acetyl-L-ornithine
N-carbamoyl-L-aspartate	citrulline	arginine
glucosamine	N-carbamoyl-L-aspartate	citrulline
tyrosine	glucosamine	N-carbamoyl-L-aspartate
Phosphorylcholine	tyrosine	glucosamine
3-phospho-serine	Phosphorylcholine	tyrosine
N6-Acetyl-L-lysine	3-phospho-serine	Phosphorylcholine
Acetyllysine	N6-Acetyl-L-lysine	3-phospho-serine
N-acetyl-glutamine	Acetyllysine	N6-Acetyl-L-lysine
N-acetyl-glutamate	N-acetyl-glutamine	Acetyllysine
Ng,NG-dimethyl-L-arginine	N-acetyl-glutamate	N-acetyl-glutamine
Acetylcarnitine DL	Ng,NG-dimethyl-L-arginine	N-acetyl-glutamate
tryptophan	Acetylcarnitine DL	Ng,NG-dimethyl-L-arginine
Kynurenine	tryptophan	Acetylcarnitine DL
N-acetyl-glucosamine	Kynurenine	tryptophan
Flavone	N-acetyl-glucosamine	Kynurenine
cystathionine	Flavone	N-acetyl-glucosamine
5-methoxytryptophan	cystathionine	Flavone
Cystine	5-methoxytryptophan	cystathionine
cytidine	Cystine	5-methoxytryptophan
biotin	cytidine	Cystine
deoxyadenosine	biotin	cytidine
Glycerophosphocholine	deoxyadenosine	biotin
acadesine	Glycerophosphocholine	deoxyadenosine
D-glucosamine-6-phosphate	acadesine	Glycerophosphocholine
D-glucosamine-1-phosphate	D-glucosamine-6-phosphate	acadesine
thiamine	D-glucosamine-1-phosphate	D-glucosamine-6-phosphate
S-ribosyl-L-homocysteine-posi	thiamine	D-glucosamine-1-phosphate
deoxyguanosine	S-ribosyl-L-homocysteine-posi	thiamine
adenosine	deoxyguanosine	S-ribosyl-L-homocysteine-posi
1-Methyladenosine	adenosine	deoxyguanosine
guanosine	1-Methyladenosine	adenosine
L-arginino-succinate	guanosine	1-Methyladenosine
S-methyl-5-thioadenosine	L-arginino-succinate	guanosine
7-methylguanosine	S-methyl-5-thioadenosine	L-arginino-succinate
dCMP	7-methylguanosine	S-methyl-5-thioadenosine
glutathione	dCMP	7-methylguanosine
dTMP	glutathione	dCMP
CMP	dTMP	glutathione
UMP	CMP	dTMP
dAMP	UMP	CMP
Nicotinamide ribotide	dAMP	UMP
aminoimidazole carboxamide ribonucleotide	Nicotinamide ribotide	dAMP

Table 4.1, continued

thiamine-phosphate	aminoimidazole carboxamide ribonucleotide	Nicotinamide ribotide
dGMP	thiamine-phosphate	aminoimidazole carboxamide ribonucleotide
AMP	dGMP	thiamine-phosphate
IMP	AMP	dGMP
S-adenosyl-L-methioninamine	IMP	AMP
GMP	S-adenosyl-L-methioninamine	IMP
xanthosine-5-phosphate	GMP	S-adenosyl-L-methioninamine
riboflavin	xanthosine-5-phosphate	GMP
S-adenosyl-L-homoCysteine-posi	riboflavin	xanthosine-5-phosphate
S-adenosyl-L-methionine	S-adenosyl-L-homoCysteine-posi	riboflavin
folate	S-adenosyl-L-methionine	S-adenosyl-L-homoCysteine-posi
7,8-dihydrofolate	folate	S-adenosyl-L-methionine
5-methyl-THF	7,8-dihydrofolate	folate
Diiodothyronine	5-methyl-THF	7,8-dihydrofolate
glutathione disulfide-posi	Diiodothyronine	5-methyl-THF
NAD+_posi	glutathione disulfide-posi	Diiodothyronine
NADH	NAD+_posi	glutathione disulfide-posi
dephospho-CoA-posi	NADH	NAD+_posi
NADP+_posi	dephospho-CoA-posi	NADH
NADPH	NADP+_posi	dephospho-CoA-posi
coenzyme A-posi	NADPH	NADP+_posi
FAD	coenzyme A-posi	NADPH
acetyl-CoA-posi	FAD	coenzyme A-posi
propionyl-CoA-posi	acetyl-CoA-posi	FAD
acetoacetyl-CoA-posi	propionyl-CoA-posi	acetyl-CoA-posi
malonyl-CoA-posi	acetoacetyl-CoA-posi	propionyl-CoA-posi
succinyl-CoA-posi	malonyl-CoA-posi	acetoacetyl-CoA-posi
	succinyl-CoA-posi	malonyl-CoA-posi
	sarcosine	succinyl-CoA-posi
		sarcosine
		putrescine
		spermidine
		spermine

Table 4.2. Clustered metabolites in Figure 4.1B

Early day cluster (green)	Late day cluster (orange)
fumarate	phosphoenolpyruvate
Maleic acid	3-phosphoglycerate
2.keto-isovalerate	Kynurenic acid
Hydroxyisocaproic acid	citrate-isocitrate
malate	citrate
p-hydroxybenzoate	deoxyribose-phosphate
a-ketoglutarate	trehalose-sucrose
2.Hydroxy.2-methylbutanedioic acid	CDP-nega
orotate	dGDP-nega
dihydroorotate	dGTP
Aminoadipic acid	NAD+_nega
Indole.3-carboxylic acid	serine
2-Isopropylmalic acid	proline
Pyrophosphate	threonine
Indoleacrylic acid	Imidazoleacetic acid
D.erythrose.4-phosphate	asparagine
ribose-phosphate	aspartate
uridine	lysine
glucose.1-phosphate	glutamine
D.sedoheptulose.1.7-phosphate	histidine
N.acetyl.glucosamine.1-phosphate	arginine
UDP.N.acetyl-glucosamine	tyrosine
betaine aldehyde	S-methyl-5-thioadenosine
choline	CMP
dimethylglycine	AMP
Creatinine	IMP
betaine	S-adenosyl-L-methionine
valine	NAD+_posi
leucine-isoleucine	
methionine	
N-acetyl-L-ornithine	
citrulline	
N-carbamoyl-L-aspartate	
tryptophan	
N-acetyl-glucosamine	
biotin	
L-arginino-succinate	
UMP	

Table 4.3. Strains used in this study[2]

Strain	Genotype	Source
Synechococcus elongatus PCC 7942	WT	ATCC
BA-000021	NSI::pPsbAI-luxCDABE/Sm^{R}[3]	This work
BA-000022	PSCB0001 KaiABC::Hyg^{R}[4]	This work
BA-000023	PSCB0002 spsA::Cm^{R}[5]	This work
BA-000024	PSCB0002 dpm::Cm^{R}	This work

Table 4.4. Plasmids used in this study[2]

Plasmid	Description	Source
V0120	Silver Standard high-copy biobrick vector	(Phillips & P. A. Silver n.d.)
pAM2314	Neutral Site I (NSI) integration vector	(Mackey et al. 2007)
pPS3230	NSI integration vector with pPsbAI-luxCDABE construct, Sm^{R}	This work
pPS3231	V0120 + KaiABC:: Hyg^{R} deletion construct	This work
pPS3232	V0120 + spsA::Cm^{R} deletion construct	This work
pPS3233	V0120 + dpm::Cm^{R} deletion construct	This work

[2] Available upon request
[3] Sm^{R}: Spectinomycin resistance cassette
[4] Hyg^{R}: Hygromycin resistance cassette
[5] Cm^{R}: Chloramphenicol resistance cassette

Table 4.5. Oligos used in this study

Oligo	Sequence
Hyg-F	AAAACTAGTTCTGTCTATTTCGTTCATCCAT
Hyg-R	AAAAAGCTTCTAAAGCCGCTAGCCC
KaiABC-KO-UP-F	GATCGATCGAATTCATGAAGATTCGTAACGTATCCATGG
KaiABC-KO-UP-R	GATCGATCACTAGTCTAGCCAGCGATCGCG
KaiABC-KO-DN-F	GATCGATCAAGCTTCGCAGTCGCTCCTGTCA
KaiABC-KO-DN-R	GATCGATCGCGGCCGCTTAAGCAGCTTCGCCAGCAG
Cm-F	AAAACTAGTGCACCTCAAAAACACCATCATACA
Cm-R	AAAAAGCTTCTGCCACCGCTGAGCAAT
SpsA-KO-UP-F	AAAGAATTCGCCCCAGAGCCAAATCC
SpsA-KO-UP-R	AAAACTAGTGCGCTAGTCAGCCTCACAG
SpsA-KO-DN-F	AAAAAGCTTCCTTTTCAGTTAAAGATTGGGC
SpsA-KO-DN-R	AAAGCGGCCGCGCGGCTTTTACAGTCGCTAC
dpm-KO-UP-F	AAAGAATTCACAGAATTGCGTCGTGTCG
dpm-KO-UP-R	AAAACTAGTATTGCTCAGAGAGCCGCTAG
dpm-KO-DN-F	AAAAAGCTTTGCACGGATGGAATCCAC
dpm-KO-DN-R	AAAGCGGCCGCCTTCGGAAGAAACGTTGCG

REFERENCES

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp.403–410.

Blumwald, E., Mehlhorn, R.J. & Packer, L., 1983. Studies of osmoregulation in salt adaptation of cyanobacteria with ESR spin-probe techniques. *Proceedings of the National Academy of Sciences of the United States of America*, 80(9), pp.2599–2602.

Cheung, K.J. et al., 2003. A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of Escherichia coli. *Genome research*, 13(2), pp.206–215.

Clerico, E.M., Ditty, J.L. & Golden, S.S., 2007. Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods in molecular biology (Clifton, NJ)*, 362, pp.155–171.

Curatti, L. et al., 1998. Sucrose-phosphate synthase from Synechocystis sp. strain PCC 6803: identification of the spsA gene and characterization of the enzyme expressed in Escherichia coli. *Journal of Bacteriology*, 180(24), pp.6776–6779.

Dalchau, N. et al., 2011. The circadian oscillator gene GIGANTEA mediates a long-term response of the Arabidopsis thaliana circadian clock to sucrose. *Proceedings of the National Academy of Sciences of the United States of America*.

Dong, G. & Golden, S.S., 2008. How a cyanobacterium tells time. *Current Opinion in Microbiology*, 11(6), pp.541–546.

Ducat, D.C., Way, J.C. & Silver, P.A., 2011. Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology*, 29(2), pp.95–103.

Golden, S.S., 2007. Integrating the circadian oscillator into the life of the cyanobacterial cell. *Cold Spring Harbor symposia on quantitative biology*, 72, pp.331–338.

Ishii, N. et al., 2007. Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science (New York, NY)*, 316(5824), pp.593–597.

Ito, H. et al., 2009. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33), pp.14168–14173.

Kelly, A.D. et al., 2011. Metabolomic profiling from formalin-fixed, paraffin-embedded tumor tissue using targeted LC/MS/MS: application in sarcoma. *PLoS ONE*, 6(10), p.e25357.

Kishony, R. & Leibler, S., 2003. Environmental stresses can alleviate the average deleterious effect of mutations. *J Biol*, 2(2), p.14.

Kondo, T. et al., 1993. Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12), pp.5672–5676.

Kucho, K.-I. et al., 2005. Global analysis of circadian expression in the cyanobacterium Synechocystis sp. strain PCC 6803. *Journal of Bacteriology*, 187(6), pp.2190–2199.

Locasale, J.W. et al., 2012. Metabolomics of Human Cerebrospinal Fluid Identifies Signatures of Malignant Glioma. *Molecular & cellular proteomics : MCP*.

Mackey, S.R., 2007. Biological Rhythms Workshop IA: molecular basis of rhythms generation. In *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor symposia on quantitative biology. pp. 7–19.

Mackey, S.R. et al., 2007. Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. *Methods in molecular biology (Clifton, NJ)*, 362, pp.115–129.

Nakajima, M. et al., 2005. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science (New York, NY)*, 308(5720), pp.414–415.

Niederholtmeyer, H. et al., 2010. Engineering synthesis and export of hydrophilic products from cyanobacteria. *Applied and Environmental Microbiology*.

Nishiwaki, T. et al., 2004. Role of KaiC phosphorylation in the circadian clock system of Synechococcus elongatus PCC 7942. *Proceedings of the National Academy of Sciences of the United States of America*, 101(38), pp.13927–13932.

Phillips, I. & Silver, P., 2006. A new biobrick assembly strategy designed for facile protein engineering. *DSpace. MIT Artificial Intelligence Laboratory: MIT Synthetic Biology Working Group, Massachusetts Institute of Technology, Cambridge, MA. http://hdl. handle. net/1721.1/32535*.

Phillips, I. & Silver, P.A., A new biobrick assembly strategy designed for facile protein engineering. *DSpace. MIT Artificial Intelligence Laboratory: MIT Synthetic Biology Working Group, Massachusetts Institute of Technology, Cambridge, MA. http://hdl. handle. net/1721.1/32535*.

Porchia, A.C. & Salerno, G.L., 1996. Sucrose biosynthesis in a prokaryotic organism: Presence of two sucrose-phosphate synthases in Anabaena with remarkable differences compared with the plant enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), pp.13600–13604.

Roenneberg, T. & Foster, R.G., 1997. Twilight times: light and the circadian system. *Photochemistry and Photobiology*, 66(5), pp.549–561.

Rust, M.J., Golden, S.S. & O'Shea, E.K., 2011. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science (New York, NY)*, 331(6014), pp.220–223.

Savage, D.F. et al., 2010. Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science (New York, NY)*, 327(5970), pp.1258–1261.

Tomita, J. et al., 2005. No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science (New York, NY)*, 307(5707), pp.251–254.

Vijayan, V., Jain, I.H. & O'Shea, E.K., 2011. A high resolution map of a cyanobacterial

transcriptome. *Genome biology*, 12(5), p.R47.

Vijayan, V., Zuzow, R. & O'Shea, E.K., 2009. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), pp.22564–22568.

Woelfle, M.A. et al., 2004. The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Current biology : CB*, 14(16), pp.1481–1486.

Wold, S., Sjöström, M. & Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory* ….

Yuan, M., Breitkopf, S. & Asara, J.M., 2012. A Positive/Negative Switching Targeted Mass Spectrometry Based Metabolomics Platform for Bodily Fluids, Cells, Fresh and Fixed Tissue. *Nat Protoc*.

Chapter 5

Conclusions

"Systems biology eats synthetic biology."

— Andy Ellington

SYNTHETIC BIOLOGY: THE NEXT GENERATION

In considering the future development of synthetic biology, progress in biological design is often compared to progress in transistor and microchip design. Most dramatically, the cost of DNA sequencing and DNA synthesis has fallen exponentially, similar to Moore's law for transistor density (Carlson 2010). This comparison often leads to the conclusion that synthetic biology will develop similarly to the computer revolution, where rapid development of off-the-shelf standardized components enabled enthusiasts to design new personal computers. While synthetic biology may someday be similarly accessible to home users, biological design is currently too unpredictable to compare to computer hardware design. Perhaps a more appropriate comparison is to the advent of powered flight, when extensive trial and error was used to reveal the basic flight principles (Carlson 2010).

Synthetic biologists hope to discover biological design principles—a set of rules that would govern the design of new biological systems—in order to eliminate the trial and error required to yield functional devices. To date, much of this effort has centered on defining and characterizing modular biological parts, developing strategies for device isolation, and chassis minimization (Boyle & Silver 2009; Boyle & Silver 2011). While this research has been quite productive, this progress has not translated into a unified philosophy for designing biology. Identifying biological design principles will allow us to design biology in a "biological" manner rather than depending on design metaphors from other disciplines.

THE BIOLOGICAL ENGINEERING DESIGN CYCLE

A complicating issue in the search for biological design principles is that natural biological systems are evolved, not designed. Yet, natural selection has yielded common sequences, folds, and motifs that appear in a variety of contexts throughout biology. Most successful applications of

synthetic biology have exploited these evolved modular parts (Agapakis & Silver 2009, Appendix F). In synthetic devices, each individual module (such as a promoter or a protein domain) often fills the role it evolved to perform, while the devices are capable of novel behavior due to the manner in which the modules are connected. Thus, evolved biological elements can be rearranged to perform emergent complex behaviors. Biological design principles will likely be informed by evolution, either through rational design that provides devices with high fitness for the desired task, or by embracing irrational design to rapidly evolve devices with the desired function (Boyle & Silver 2011).

While we currently lack fundamental design principles, engineering design cycles may improve our ability to predictably engineer biology. There are two essential properties of an engineering design cycle (Figure 5.1). First, the test portion of the cycle must identify designs that fail to perform the desired task. Second, the results of the test phase must facilitate improvement of the model for the next turn of the cycle. Ideally, an engineering design cycle fails gracefully: unexpected results from each turn of the cycle help improve the model. Ultimately, such a cycle may reveal biological design principles.

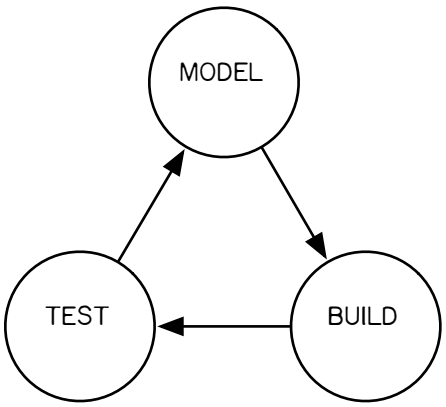


Figure 5.1 The ideal engineering design cycle. Models for the *a priori* design of biological systems do not necessarily have to be accurate. Instead, imperfect models that can be refined by experimental results can facilitate the iterative development of improved biological systems.

MODELING AND TESTING AT THE NETWORK SCALE

A confounding factor in effectively integrating the model and test phases is that a model's predictions are often difficult to experimentally verify, and vice versa. In the case of metabolic engineering, researchers are primarily concerned with the conversion efficiency from feedstock to product. However, as demonstrated in this dissertation and other metabolic engineering efforts, understanding internal metabolic fluxes is important to improving pathway throughput (Chapter 2, Fell 1997; Solomon & Prather 2011). Flux balance analysis and other constraint-based metabolic models estimate metabolic flux values, but metabolic fluxes themselves are difficult to measure directly. Unlike the genome and the transcriptome, which are composed of nucleic acid polymers, the metabolome is a complex mixture of small molecules with disparate chemical properties. As a result, mass spectrometry approaches that measure 13C flux are time consuming and only measure a subset of the metabolic network (Yuan et al. 2008; Bennett et al. 2009). These technical limitations could potentially be resolved by developing frameworks to translate data between networks that can be modeled and networks that can be measured (Figure 5.2).

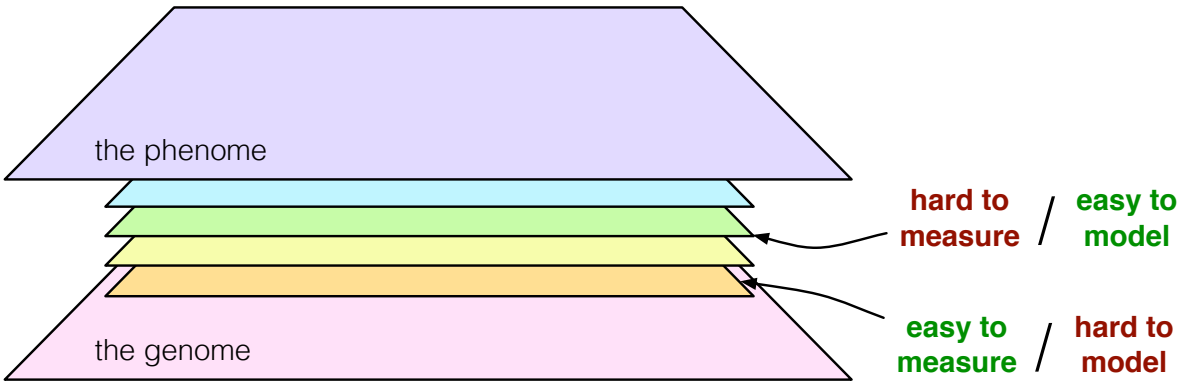


Figure 5.2 Design trade-offs in the meta-ome. Between the genome (the –ome that we can directly manipulate) and the phenome (the –ome that we aim to influence) lie several layers of regulation that impact cell behavior. The distinct properties of each layer affect the biological design cycle. The transcriptome (orange) is easy to measure, but transcript levels are difficult to relate to metabolite concentrations. The metabolome (green) is more difficult to measure, but constraint-based genome scale models can make *a priori* predictions of the entire metabolic network.

Synthetic biologists seek to control the phenome of biological systems; this is accomplished by engineering the genome of these systems. Between these two layers are several layers of regulation, each with unique properties that define our ability observe and model that layer (Figure 5.2). As described in Chapter 1, several studies have explored the integration of data from the transcriptome layer with network-scale data from the metabolome layer (Bradley et al. 2009; Fendt et al. 2010; Ishii et al. 2007; Moxley et al. 2009). Although this research is in its early stages, the work done thus far suggests that transcriptome data is insufficient to reconstruct the status of the metabolome. Our work in Chapter 4 observed metabolic dynamics in *S. elongatus* under control of the kai circadian clock. This system was chosen in part because the transcriptome of *S. elongatus* is globally regulated by the kai clock, thus it would be reasonable to expect that circadian oscillations would likewise be prevalent in the metabolome. Instead, we observed that other factors, such as the presence of light, had an equal or greater effect on metabolome status. Similarly, other studies have observed that metabolite concentrations often remain constant in response to enzyme perturbations (Ishii et al. 2007; Fendt et al. 2010), and that transcript-metabolite correlations are dependent on factors such as media composition and the type of metabolite (Bradley et al. 2009).

Given the mixed results observed thus far, it is clear that simple correlation of transcriptome and metabolome data is unlikely to yield universal insight into the regulation of metabolism (Bradley et al. 2009). Instead, global comparisons of the transcriptome and metabolome at the network scale can inform more detailed analysis at the pathway level. For example, transcriptome analysis of our formic acid producing yeast revealed coordinated regulation of the serine-glycine biosynthesis pathway. Based on this analysis, we identified 3-phosphoglycerate dehydrogenase as having a high degree of flux control over this pathway, and observed that overexpressing this enzyme improved formic acid yields (Chapter 2). Similarly, the circadian regulation of sucrose production that we observed in cyanobacteria would have been difficult to identify *a priori* from transcriptome data. In-

depth analysis of the sucrose pathway could potentially reveal how global oscillations in transcription impact local pathway dynamics (Chapter 4). Detailed kinetic approaches to studying metabolism, such as Metabolic Control Analysis (Fell 1997), are too complex to model the entire metabolome (Schuster et al. 1999), but are ideally suited for local pathway modeling. Thus, integrating small-scale and network-scale approaches enhances engineering design cycles.

SYSTEMS SYNTHETIC BIOLOGY

In an effort to facilitate the development of biological engineering design cycles, this dissertation has presented systems approaches to modeling and testing biological designs. Synthetic biology resembles systems biology in reverse: quantitative models of biological components are used to build new interactions rather than to understand endogenous interactions. Synthetic biology often differs from systems biology in the consideration of network-scale biological systems: synthetic biologists commonly seek to build devices that are orthogonal or otherwise insulated from the host chassis. While this approach is valuable for improving the robustness of many synthetic devices, including the hydrogenase-based circuits we constructed in Chapter 3 of this work, it is not feasible for many metabolic engineering objectives. As demonstrated in Chapter 2, re-routing central metabolic flux towards a desired product is a systems problem, involving the entire metabolic network and associated regulatory networks.

Network-scale models and measurements provide insight into unanticipated interactions between synthetic devices and host systems. In Chapter 2, our model identified a non-intuitive set of gene knockouts that improved formic acid production; the enzymes targeted were not closely associated with the serine-glycine biosynthesis pathway that produces formic acid in yeast. Also in Chapter 2, systems-level analysis of transcription changes identified unforeseeable regulatory responses to our gene knockouts. Together, these results underscore the value of design cycles: the

model identified a strain unlikely to have been isolated via experimental screening, and our subsequent experiments identified strain improvements that could not have been discovered by our model.

Network-scale effects also complicate efforts to integrate modular parts and devices into new hosts. In Chapter 3, for example, we deleted genes to increase carbon flux to the hydrogenase circuits and to eliminate unwanted ferredoxin-mediated side reactions that would siphon reducing equivalents away from the circuits. Selecting genes to knock out was done primarily in an ad hoc manner, via literature searches and bioinformatics. This process was time-consuming, and many of the knockouts had no effect on phenotype or were deleterious. This experience underscores the importance of part characterization, even in well-studied hosts such as *Escherichia coli*. Once again, systems biology approaches will be essential: we must study how host networks impact device function, not just individual proteins.

Conversely, we must also study how devices impact network behavior. The Kai clock is a robust modular oscillator that exhibits circadian phosphorylation oscillations *in vitro* (Nakajima et al. 2005), and *in vivo* these oscillations control most of the transcriptional network of *Synechococcus elongatus* (Ito et al. 2009; Vijayan et al. 2009). Systems biology analysis of this system is important to establishing synthetic biology design principles, as it is a natural example of what we aim to engineer: network-scale control of host systems via a well-characterized modular device.

Given recent successes in DNA synthesis, DNA sequencing, and DNA assembly, the "build" phase of the synthetic biology engineering design cycle is no longer limiting (Boyle & Silver 2009). What remains are tasks well suited to systems biology: network-scale modeling and network scale measurement of biological systems. Both modeling and measurement are dependent on each other: models provide potential explanations for measurements, and measurements identify oversights in models. As demonstrated in this dissertation, a combined modeling and measurement

approach is an integral part of establishing biological design cycles. The future of synthetic biology as an engineering discipline will be determined by design cycles that fully integrate the model, build, and test phases.

REFERENCES

Agapakis, C.M. & Silver, P.A., 2009. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. *Molecular BioSystems*, 5(7), pp.704–713.

Bennett, B.D. et al., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology*, 5(8), pp.593–599.

Boyle, P.M. & Silver, P.A., 2009. Harnessing nature's toolbox: regulatory elements for synthetic biology. *Journal of the Royal Society, Interface / the Royal Society*, 6 Suppl 4, pp.S535–46.

Boyle, P.M. & Silver, P.A., 2011. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metabolic Engineering*.

Bradley, P.H. et al., 2009. Coordinated concentration changes of transcripts and metabolites in Saccharomyces cerevisiae. *PLoS Computational Biology*, 5(1), p.e1000270.

Carlson, R.H., 2010. *Biology is Technology: The Promise, Peril, and New Business of Engineering Life*, Cambridge, MA: Harvard University Press.

Fell, D., 1997. *Understanding the Control of Metabolism* K. Snell, ed., London: Portland Press.

Fendt, S.-M. et al., 2010. Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Molecular Systems Biology*, 6, p.356.

Ishii, N. et al., 2007. Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science (New York, NY)*, 316(5824), pp.593–597.

Ito, H. et al., 2009. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33), pp.14168–14173.

Moxley, J.F. et al., 2009. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16), pp.6477–6482.

Nakajima, M. et al., 2005. Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science (New York, NY)*, 308(5720), pp.414–415.

Schuster, S., Dandekar, T. & Fell, D.A., 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology*, 17(2), pp.53–60.

Solomon, K.V. & Prather, K.L.J., 2011. The zero-sum game of pathway optimization: emerging paradigms for tuning gene expression. *Biotechnology Journal*, 6(9), pp.1064–1070.

Vijayan, V., Zuzow, R. & O'Shea, E.K., 2009. Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52), pp.22564–22568.

Yuan, J., Bennett, B.D. & Rabinowitz, J.D., 2008. Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nature Protocols*, 3(8), pp.1328–1340.

Appendix A

Supporting Information for Chapter 3[1]

[1] Portions of the work presented in this chapter were published in the following papers:

1. Agapakis, C. M., Ducat, D.C., Boyle, P.M., Wintermute, E.H., Way, J.C., & Silver, P.A. Insulation of a synthetic hydrogen metabolism circuit in bacteria. J Biol Eng **4**, 3 (2010).
2. Barstow, B., Agapakis, C. M.*, Boyle, P.M.*, Gerald, G.*, Silver, P.A., & Wintermute, E.H. A synthetic system links FeFe-hydrogenases to essential E. coli sulfur metabolism. *J Biol Eng* **5**, 7–7 (2011). *Equal contribution.

Supplementary Information for Agapakis et. al. "Insulation of a synthetic hydrogen metabolism circuit in bacteria"

>Cr.HydA1;AAL23572
gaattcgcggccgcttctagagctgcaccagccgcagaagctcctttgtctcatgttcaacaggccttag
ccgagcttgcaaaaccaaaggatgaccctactagaaaacacgtatgtgtccaagtggccccagctgttag
ggtagcaattgctgaaacacttggtttggccccctggagcaaccactccaaagcagttagctgagggccta
agaaggcttggttttgatgaagtgttcgacacattgtttggagccgatttaaccataatggaagagggct
cagaattgttacatagactaactgaacaccttgaggcacatcctcactccgacgaaccattgcctatgtt
cacaagttgctgtccaggttggatcgctatgttagaaaaaagctatcctgatctaattccatacgtgagc
tcatgcaagtcccctcaaatgatgttggccgcaatggttaaaagttatttagctgagaagaaaggtatag
ccccaaaggatatggtaatggtcagcatcatgccatgtaccagaaaacaatctgaagcagacagggattg
gttttgcgttgacgctgatcctactcttagacagttggatcatgtgattacaaccgttgagttaggaaat
atattcaaggaaagaggcatcaacctagccgaacttccagagggtgaatgggacaatcctatgggagtag
gttcaggcgcaggtgtcttgtttggaactacaggcggcgtgatggaagctgctttaaggactgcctacga
gctattcaccggtacaccattgcctagattatcccttagtgaagttaggggaatggatggtattaaagaa
actaacattaccatggtaccagcacctggctctaagtttgaggaattgttaaaacatagagctgccgcaa
gagctgaagccgcagctcacggaacaccaggtcctctagcatgggacggcggtgctggattcactagcga
ggatggtaggggcggcataacattgagagtcgccgttgcaaatggattaggtaacgctaaaaagcttatc
accaaaatgcaagccggcgaagcaaagtatgattttgtggagattatggcttgtccagccggatgtgttg
gtggaggcggacaacctagatcaactgacaaagcaataacacagaagaggcaagctgccctatacaattt
ggatgaaaaatccactttaagaagaagtcatgaaaacccatctatcagggagctttatgacacctacttg
ggtgaacctttaggtcacaaggcacatgaactattgcacacacattatgtagctggcggagtcgaggaaa
aagatgaaaagaaaactagtagcggccgctgcag

>Cr.HydEF;AAS92601
gaattcgcggccgcttctagagctgcacatgcctctgcttcaaaagcaactccagatgttcctgtagacg
atcttccacctgcccacgctagagcagccgtcgccgcagctaataggagagccagggcaatggcttccgc
cgaagcagctgccgagacattaggtgactttctaggacttggcaagggtggattgagtccaggcgcaacc
gctaacttagatagagaacaagtgctaggtgttcttgaggccgtatggagaaggggtgacttgaatttag
aaagagcattgtatagccatgctaacgccgtcactaataaatactgtggaggcggtgtgtattacagagg
attagttgagttctctaacatttgccagaatgattgttcatattgcggtataaggaacaatcaaaaggag
gtatggagatacacaatgcctgtcgaagaagttgtggaggttgcaaaatgggccctagaaaacggcatca
ggaatattatgcttcagggtggagaacttaagaccgagcaaagattagcttacctagaagcctgtgtaag
agcaataagggaggaaactacacaattggatttagaaatgagagctagagccgcatccaccactacagct
gaggccgcagctagtgcacaggctgacgccgaagcaaaaaggggtgaaccagagcttggcgtcgtggtta
gcttgtctgtaggtgaattacctatggaacaatacgagagactatttagagctggagccaggagatatct
tatcaggattgaaacctcaaatccagatttgtacgcagctttacaccctgaaccaatgtcctggcatgcc
agagtcgagtgcctaagaaacttgaagaaagcaggttatatgttaggcactggagttatggtgggccttc
ctggccaaacattgcacgacttagctggtgatgttatgttctttagggatataaaggccgacatgatcgg
aatgggtccattcattactcagcctggcaccccagcaacagataaatggactgctctataccccaaatgct
aacaagaatagtcatatgaaatctatgtttgacttgaccacagccatgaacgcattagtaagaattacta
tgggtaatgtcaacataagcgctacaaccgcccttcaagcaatcattcctactggaagagaaatagccct
agagaggggtgccaatgtggttatgccaatcttgacacctactcagtatagagaatcataccaattatat
gaaggcaagccatgtattaccgatacagcagtacaatgtagaaggtgccttgatatgagattgcattccg
tcggaaaaaccagtgctgccggtgtttggggtgaccctgcatctttcttacacccaatagtgggcgttcc
tgtaccacatgatctatcatctcctgctttggccgcagctgccagcgcagactttcacgaggtcggagct
ggtccatggaaccctatcaggttagaaagacttgttgaagtgccagatagataccctgatccagacaatc
atggtaggaaaaaggccggcgcaggaaaaggcggcaaggctcacgattcccatgacgatggagatcatga
cgatcaccatcaccatcatggtgccgcaccagctggtgccgcagctggcaaaggaaccggtgccgcagct
attggcggcggagccggtgctagcagacagagagtagctggcgccgcagctgcctcagcaaggttgtgtg
ctggagccagaagagcaggtagggtcgttgcttctcctctaagaccagccgcagcttgcaggggtgtggc
cgttaaggcagctgctgccgcagctggcgaggacgccggagcaggtacaagcggtgtaggctccaatatt
gtcaccagtcctggaatagcttcaaccacagcccacggtgttccaagaatcaacattggcgtgttcggag
taatgaatgcaggtaaatctactttagtcaacgctttggcccaacaagaagcatgtatagttgatagcac
ccctggtacaactgctgacgtcaagaccgttcttttagaactacatgcattgggcccagctaaattactt

```
gatacagccggattggatgaggtaggtggtctaggcgacaagaaaagaaggaaggcattaaatactttga
aagaatgcgatgtcgctgttcttgtggtagacaccgatacagccgcagctgccatcaaatccggaagatt
agcagaggccctagaatgggaaagtaaggtcatggagcaggctcacaaatacaacgtttcacctgtgttg
ttattgaatgtaaagagcagaggccttccagaagcccaagcagccagcatgctagaagccgttgcaggca
tgttagatccttccaaacagattccaaggatgtcattggacttagcttctactcctcttcatgagagaag
tacaataactagcgcctttgtcaaggaaggagcagttaggtcctcaagatacggtgctccactacctggt
tgtttgccaagatggtctttaggcaggaacgccagattgcttatggtgattccaatggatgcagaaaccc
ctggaggtagactattaaggccacaagctcaagtaatggaggaagccatcagacactgggcaacagtctt
gagtgttagattagacttggatgctgccaggggtaaacttggccctgaagcatgtgagatggaaagacag
aggttcgatggagtaattgctatgatggagagaaatgacggtccaactctagttgtgaccgattctcaag
ccatagacgtcgttcatccttggacattagatagatcctcaggcaggccattggtgcctatcactacctt
tagtattgcaatggcttatcaacagaacggaggtagacttgatccatttgtagaaggcctagaagcctta
gagacattgcaagacggcgatagagtcttaatatctgaagcatgcaatcataataggatcacttcagctt
gtaacgacattggaatggttcaaatacctaataagttggaagctgcacttggtggtaaaaagctacagat
tgagcacgctttcggcagagaatttccagaattagagtctggaggtatggatggcttgaaacttgccatc
cattgcggaggttgtatgattgatgcacaaaagatgcagcaaagaatgaaagacctacacgaagctggtg
tacctgttaccaactatggcgtgttctttagctgggccgcatggccagatgctttaaggagagccttgga
accttggggagtcgagcctccagttggtacacctgcaactccagctgccgcacctgctaccgccgcatcc
ggtgtgactagtagcggccgctgcag
```

```
>Cr.HydG;AAS92602
gcggccgcttctagaactgctcatggtaaagcatctgccacaagagaatatgctggagatttttgccag
gcaccactatttcacacgcatggtccgttgagagggaaacacatcacagatacaggaatcctgccgagtg
gataaacgaagctgcaatccataaggccttagaaaccagtaaagctgacgcacaagatgctggtagagta
agagagattctagccaaggcaaaagaaaaggctttcgtcactgaacacgccccagtgaatgcagagagca
aatctgaatttgttcagggacttacattggaagagtgtgctaccttaataaacgtagactcaaataacgt
cgaactaatgaatgagatcttcgatactgcccttgcaattaaggaaaggatatatggcaacagagtggtt
ttgtttgctcctttatacatcgccaatcattgcatgaacacatgtacctattgcgcattcagatccgcta
ataaaggtatggaaaggagtattttgactgacgatgatttaagagaggaagtagccgcactacaaaggca
gggtcatagaaggattcttgctttgacaggagaacacccaaagtacacttttgacaatttcttacatgct
gtcaacgttatagccagcgtgaaaaccgagcctgaaggctctatcaggagaattaatgttgaaatcccac
ctctatcagtatccgatatgagaaggttgaagaacacagacagtgtcggtacttttgtgttattccaaga
gacctatcacagagatacatttaaagttatgcatccatctggacctaagagcgatttcgactttagagta
cttactcaagatagggcaatgagagctggtttggacgatgtcggcatcggtgccttatttggactatacg
attataggtacgaagtttgtgcaatgcttatgcactcagaacatttggagagagaatataatgctggtcc
acatacaatttccgtgcctagaatgaggccagccgacggcagtgagttatctatagcacctccataccca
gttaacgatgctgacttcatgaagctagtagcagtcttgagaatcgctgtgccttataccggtatgattt
tatcaactagagaatctccagaaatgaggagcgccctttgaaatgcggaatgtcccagatgagtgcagg
ttcaagaacagatgttggcgcttaccacaaggatcatactttatctaccgaggccaatctaagcaaattg
gcaggacaatttacattacaagacgaaagacctactaacgaaattgtaaagtggcttatggaggaaggtt
atgtcccatcctggtgtaccgcttgttacaggcagggcagaacaggtgaagatttcatgaatatatgcaa
agccggagacatccacgatttttgtcatcctaacagtctattgactttacaagagtatcttatggattac
gcagacccagatttgaggaagaaaggtgaacaggttattgctagagagatgggccctgacgcctcagaac
cattatctgcacaaagcagaaagaggctagaaagaaaaatgaagcaagtgttggagggtgaacatgatgt
ttatttaactagtagcggccg
```

```
>So.Fd;1704156A
gcggccgcttctagagctgcatataaagttactttggtaacaccaaccggtaatgtcgaatttcaatgtc
ctgatgacgtgtacattttagacgccgctgaggaagagggaatagatctaccatattcttgcagagcagg
ctcatgttccagttgcgccggtaagcttaaaactggaagcttgaaccaggatgaccaatctttcttagat
gatgaccagatcgatgaaggctgggttctaacatgtgctgcataccctgtatcagacgtcaccattgaaa
ctcataaggaggaagaacttacagccactagtagcggccg
```

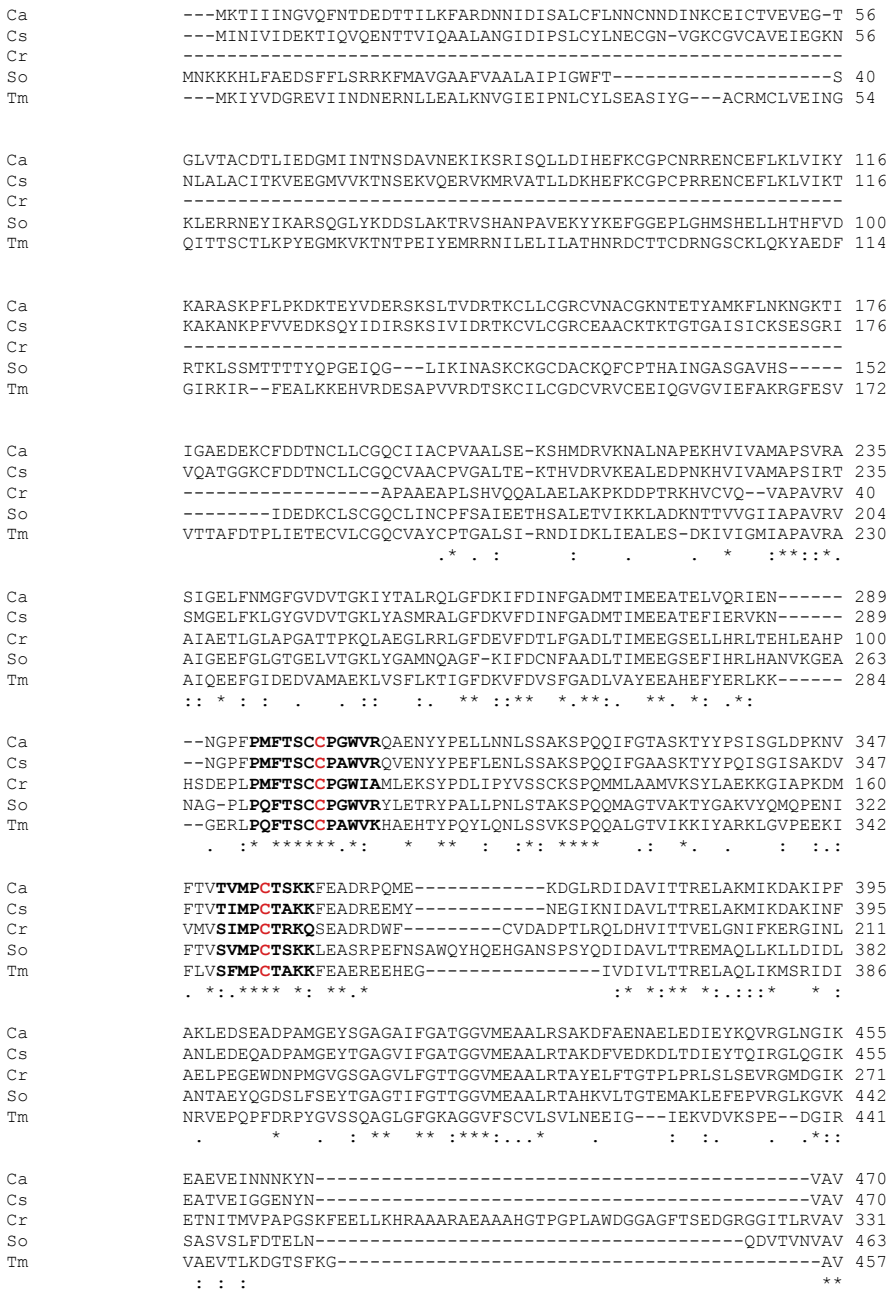
Figure S1. Sequence of codon optimized commercially synthesized genes. Gene names listed as Organism.Gene Name; GenBank Accession Number (Cr=*Chlamydomonas reinhardtii*, So=*Spinacia olearcea*)

Gene	GenBank Accession	Primer (Italics indicate Biobrick sites)
Ca HydA 5'	CAC3230	*CCTTTCTAGA* ATGAAAACAATAATCTTAAA
Ca HydA 3'		*AAGGCTGCAGCGGCCGCTACTAGT* TTCATGTTTTGAAACATTTTTA
Cs HydA 5'	CAU09760	*CCTTGAATTCGCGGCCGCATCTAGA* ATGATAAACATAGTAATTGATGAAA
Cs HydA 3'		*AAGGCTGCAGCGGCCGCTACTAGT* TTTATTGTATTTTAAGTGTAATAAT
So HydA 5'	SO_3920	*CCTTGAATTCGCGGCCGCATCTAGA* ATGACAACGACAACTTATCAACCAG
So HydA 3'		*AAGGCTGCAGCGGCCGCTACTAGT* AATGTTACCCAGCCATGAAGA GCCT
So HydB 5'	SO_3921	*CCTTGAATTCGCGGCCGCATCTAGA* ATGAACAAGAAAAAACACCTATTTG
So HydB 3'		*AAGGCTGCAGCGGCCGCTACTAGT* AGAGCTTAATTTGGTGCGATCGACA
Tm HydA 5'	AF044577	*CCTTGAATTCGCGGCCGCATCTAGA* ATGAAAATTTACGTTGATGGAAGAGAAGTTAT
Tm HydA 3'		*AAGGCTGCAGCGGCCGCTACTAGT* TCAGCCATTTTTCGAAAGCTCCTCCAGCACCTTCTC
Ca PFOR 5'	CAC2229	*CCTTGAATTCGCGGCCGCATCTAGA* ATGAAAAAAATGAAAACTATGG
Ca PFOR 3'		*AAGGCTGCAGCGGCCGCTACTAGT* TTATTGATTAGCTAATCTTT
Da PFOR 5'	CAA70873	AAACATATGGGAAAGAAAATGATGACG
Da PFOR 3'		AAACCTAGGCTACTTCTTCGTCCGCTTGC
Ec ydbK 5'	CAQ31878	*CCTTGAATTCGCGGCCGCATCTAGA* ATTACTATTGACGGTAATGGCG
Ec ydbK 3'		*AAGGCTGCAGCGGCCGCTACTAGT* TTAATCGGTGTTGCTTTTTTCCGC
Ca Fd 5'	CAC0303	*CCTTGAATTCGCGGCCGCATCTAGA* ATGGCATATAAAATAACAGA
Ca Fd 3'		*AAGGCTGCAGCGGCCGCTACTAGT* CTCTTGAACTGGAGCTCCTA
Zm Fd 5'	EU956042	*CCTTGAATTCGCGGCCGCATCTAGA* ATGGCTGTATACAAGGTGAAGCTTG
Zm Fd 3'		*AAGGCTGCAGCGGCCGCTACTAGT* CAGGTCGCCTTCCTTGTGGGTGTGG

Ca mut. 5'	GTGGACAATGTTCCAGAAGAGAAAATTGTGAGTTCCTTAAACTTG
Cs mut. 5'	GCAGCAAACTTAGCTGAGTTCATGAATAGCGG
SoB mut.	CGCGGTAGAAAAGTACTACAAAGAGTTCGGTGGCGAGCCATTAGGACATATGTCCC
CrHydA P2K	CCGCTTCTAGAGCTGCAAAGGCCGCAGAAGCTCCTTTGTC
CrHydA D126K	CGCTATGTTAGAAAAAAGCTATCCTAAGCTAATTCCATACGTGAGC
CrHydA E5K	CCGCTTCTAGAGCTGCACCAGCCGCAAAGGCTCCTTTGTC
CrHydA M119K	CCAGGTTGGATCGCTAAGTTAGAAAAAAGCTATCC

Table S1: Primers used for cloning and mutagenesis of heterologous pathway components

Figure S2. Sequence of modified multiple cloning sites of Novagen Duet Vectors. 5' phosphorylated oligonucleotide inserts (Integrated DNA Technologies, red text) were inserted between *Nco* I and *Afl* II sites in MCS1 and *Nde* I and *Avr* II of MCS2 of Novagen Duet Vectors pET-Duet, pACY-duet, pCDF-duet, and pCOLA-duet for heterologous expression of up to eight BioBrick sequences


Ca ---MKTIIINGVQFNTDEDTTILKFARDNNIDISALCFLNNCNNDINKCEICTVEVEG-T 56
Cs ---MINIVIDEKTIQVQENTTVIQAALANGIDIPSLCYLNECGN-VGKCGVCAVEIEGKN 56
Cr --
So MNKKKHLFAEDSFFLSRRKFMAVGAAFVAALAIPIGWFT--------------------S 40
Tm ---MKIYVDGREVIINDNERNLLEALKNVGIEIPNLCYLSEASIYG---ACRMCLVEING 54
Ca GLVTACDTLIEDGMIINTNSDAVNEKIKSRISQLLDIHEFKCGPCNRRENCEFLKLVIKY 116
Cs NLALACITKVEEGMVVKTNSEKVQERVKMRVATLLDKHEFKCGPCPRRENCEFLKLVIKT 116
Cr --
So KLERRNEYIKARSQGLYKDDSLAKTRVSHANPAVEKYYKEFGGEPLGHMSHELLHTHFVD 100
Tm QITTSCTLKPYEGMKVKTNTPEIYEMRRNILELILATHNRDCTTCDRNGSCKLQKYAEDF 114
Ca KARASKPFLPKDKTEYVDERSKSLTVDRTKCLLCGRCVNACGKNTETYAMKFLNKNGKTI 176
Cs KAKANKPFVVEDKSQYIDIRSKSIVIDRTKCVLCGRCEAACKTKTGTGAISICKSESGRI 176
Cr --
So RTKLSSMTTTTYQPGEIQG---LIKINASKCKGCDACKQFCPTHAINGASGAVHS----- 152
Tm GIRKIR--FEALKKEHVRDESAPVVRDTSKCILCGDCVRVCEEIQGVGVIEFAKRGFESV 172
Ca IGAEDEKCFDDTNCLLCGQCIIACPVAALSE-KSHMDRVKNALNAPEKHVIVAMAPSVRA 235
Cs VQATGGKCFDDTNCLLCGQCVAACPVGALTE-KTHVDRVKEALEDPNKHVIVAMAPSIRT 235
Cr ------------------APAAEAPLSHVQQALAELAKPKDDPTRKHVCVQ--VAPAVRV 40
So --------IDEDKCLSCGQCLINCPFSAIEETHSALETVIKKLADKNTTVVGIIAPAVRV 204
Tm VTTAFDTPLIETECVLCGQCVAYCPTGALSI-RNDIDKLIEALES-DKIVIGMIAPAVRA 230
.* . : : . . * :**::*.
Ca SIGELFNMGFGVDVTGKIYTALRQLGFDKIFDINFGADMTIMEEATELVQRIEN------ 289
Cs SMGELFKLGYGVDVTGKLYASMRALGFDKVFDINFGADMTIMEEATEFIERVKN------ 289
Cr AIAETLGLAPGATTPKQLAEGLRRLGFDEVFDTLFGADLTIMEEGSELLHRLTEHLEAHP 100
So AIGEEFGLGTGELVTGKLYGAMNQAGF-KIFDCNFAADLTIMEEGSEFIHRLHANVKGEA 263
Tm AIQEEFGIDEDVAMAEKLVSFLKTIGFDKVFDVSFGADLVAYEEAHEFYERLKK------ 284
:: * : : . . :: :. ** ::** *.**:. **. *: .*:
Ca --NGPFPMFTSCCPGWVRQAENYYPELLNNLSSAKSPQQIFGTASKTYYPSISGLDPKNV 347
Cs --NGPFPMFTSCCPAWVRQVENYYPEFLENLSSAKSPQQIFGAASKTYYPQISGISAKDV 347
Cr HSDEPLPMFTSCCPGWIAMLEKSYPDLIPYVSSCKSPQMMLAAMVKSYLAEKKGIAPKDM 160
So NAG-PLPQFTSCCPGWVRYLETRYPALLPNLSTAKSPQQMAGTVAKTYGAKVYQMQPENI 322
Tm --GERLPQFTSCCPAWVKHAEHTYPQYLQNLSSVKSPQQALGTVIKKIYARKLGVPEEKI 342
. :* ******.*: * ** : :*: **** .: *. . : :.:
Ca FTVTVMPCTSKKFEADRPQME------------KDGLRDIDAVITTRELAKMIKDAKIPF 395
Cs FTVTIMPCTAKKFEADREEMY------------NEGIKNIDAVLTTRELAKMIKDAKINF 395
Cr VMVSIMPCTRKQSEADRDWF---------CVDADPTLRQLDHVITTVELGNIFKERGINL 211
So FTVSVMPCTSKKLEASRPEFNSAWQYHQEHGANSPSYQDIDAVLTTREMAQLLKLLDIDL 382
Tm FLVSFMPCTAKKFEAEREEHEG----------------IVDIVLTTRELAQLIKMSRIDI 386
. *:.**** *: **.* :* *:** *:.:::* * :
Ca AKLEDSEADPAMGEYSGAGAIFGATGGVMEAALRSAKDFAENAELEDIEYKQVRGLNGIK 455
Cs ANLEDEQADPAMGEYTGAGVIFGATGGVMEAALRTAKDFVEDKDLTDIEYTQIRGLQGIK 455
Cr AELPEGEWDNPMGVGSGAGVLFGTTGGVMEAALRTAYELFTGTPLPRLSLSEVRGMDGIK 271
So ANTAEYQGDSLFSEYTGAGTIFGTTGGVMEAALRTAHKVLTGTEMAKLEFEPVRGLKGVK 442
Tm NRVEPQPFDRPYGVSSQAGLGFGKAGGVFSCVLSVLNEEIG---IEKVDVKSPE--DGIR 441
. * . : ** ** :***:...* . : :. . .*::
Ca EAEVEINNNKYN---VAV 470
Cs EATVEIGGENYN---VAV 470
Cr ETNITMVPAPGSKFEELLKHRAAARAEAAAHGTPGPLAWDGGAGFTSEDGRGGITLRVAV 331
So SASVSLFDTELN---QDVTVNVAV 463
Tm VAEVTLKDGTSFKG---AV 457
: : : **

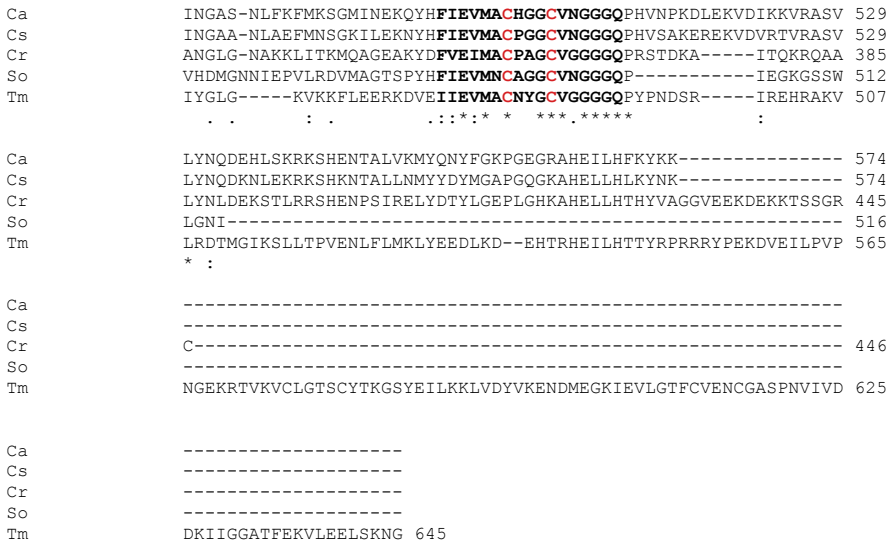


Figure S3. Sequence alignment of five hydrogenases Protein sequences of *Clostridium acetobutylicum* (Ca), *Clostridium saccharobutylicum* (Cs), *Chlamydomonas reinhardtii* (Cr), and *Thermotoga maritima (Tm)* HydA and *Shewanella oneidensis* HydB + HydA aligned using ClustalW web server (http://www.ebi.ac.uk/Tools/clustalw/). Catalytic site binding area highlighted in bold with critical cysteine residues in red.

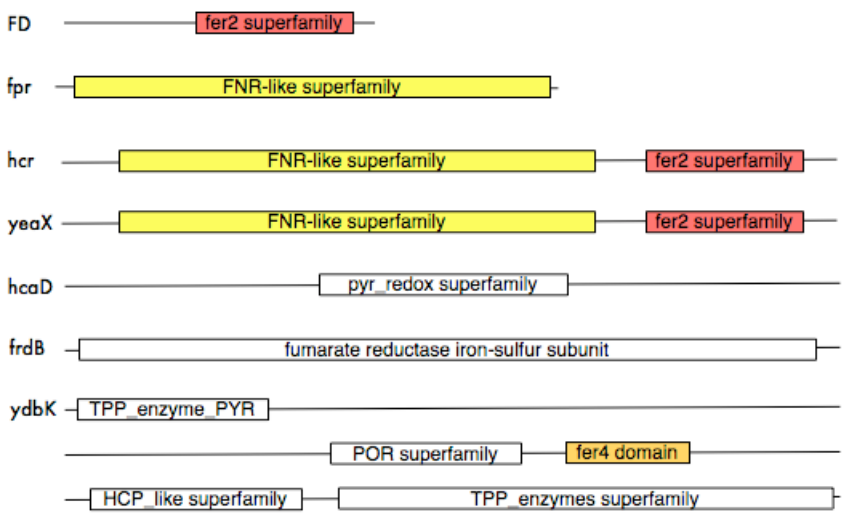


Figure S4. Deletion of competing reactions leads to hydrogen circuit insulation. A.) Domain structure of deleted ferredoxin-homology genes. FD-ferredoxin; fpr-flavodoxin:$NADP^+$ reductase; hcr-NADH oxidoreductase; yeaX-predicted oxidoreductase; hcaD-ferredoxin:NAD^+ reductase; frdB-fumarate reductase; ydbK-putative pyruvate-ferredoxin oxidoreductase. Genes were identified by BLAST homology search of the *Escherichia coli* genome against *Spinacia olearcea* ferredoxin I. Domain structure schematized from NCBI conserved domain search.

```
HydA1           APAAEAPLSHVQQALAELAKPKDDPTRKHVCVQVAPAVRVAIAETLGLAPGATTPKQLAE 60
HydA2           -ATATDAVPHWKLALEELDKPKDG-GRKVLIAQVAPAVRVAIAESFGLAPGAVSPGKLAT 58
                 .:*  .:.* : ** ** ****.  ** : .************::******.:* :**

HydA1           GLRRLGFDEVFDTLFGADLTIMEEGSELLHRLTEHLEAHPHSDEPLPMFTSCCPGWIAML 120
HydA2           GLRALGFDQVFDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPLPMFTSCCPGWVAMM 118
                *** ****:******.*********:******.************************:**:

HydA1           EKSYPDLIPYVSSCKSPQMMLAAMVKSYLAEKKGIAPKDMVMVSIMPCTRKQSEADRDWF 180
HydA2           EKSYPELIPFVSSCKSPQMMMGAMVKTYLSEKQGIPAKDIVMVSVMPCVRKQGEADREWF 178
                *****:***:**********:.****:**:**:**..**:****:***.***.****:**

HydA1           CVDADPTLRQLDHVITTVELGNIFKERGINLAELPEGEWDNPMGVGSGAGVLFGTTGGVM 240
HydA2           CVS-EPGVRDVDHVITTAELGNIFKERGINLPELPDSDWDQPLGLGSGAGVLFGTTGGVM 237
                **. :* :*::******.*************.***:.:**:*:*:*************** 

HydA1           EAALRTAYELFTGTPLPRLSLSEVRGMDGIKETNITMVPAPGSKFEELLKHR-------- 292
HydA2           EAALRTAYEIVTKEPLPRLNLSEVRGLDGIKEASVTLVPAPGSKFAELVAERLAHKVEEA 297
                *********:.*  *****.******:*****:.:*:******** **: .*

HydA1           AAARAEAAAHGTPG-PLAWDGGAGFTSEDGRGGITLRVAVANGLGNAKKLITKMQAGEAK 351
HydA2           AAAEAAAAVEGAVKPPIAYDGGQGFSTDDGKGGLKLRVAVANGLGNAKKLIGKMVSGEAK 357
                ***.* **..*:   *:*:*** **:::**:**:.*************** ** :****

HydA1           YDFVEIMACPAGCVGGGGQPRSTDKAITQKRQAALYNLDEKSTLRRSHENPSIRELYDTY 411
HydA2           YDFVEIMACPAGCVGGGGQPRSTDKQITQKRQAALYDLDERNTLRRSHENEAVNQLYKEF 417
                ************************ **********:***:.******** ::.:**. :

HydA1           LGEPLGHKAHELLHTHYVAGGVEEKDEKKTSSGRC 446
HydA2           LGEPLSHRAHELLHTHYVPGGAEADA--------- 443
                *****.*:**********.**.* .
```

Figure S5. Alignment of *Chlamydomonas reinhardtii* HydA1 and HydA2. ClustalW alignment of HydA1 and HydA2 with mutations predicted to improve the binding between HydA2 and ferredoxin (Long et. al., 2009) highlighted in red.

Table 1 Plasmids used in this study

Name	Constructs		Backbone	Resistance	Source
Hydrogenase activity *in vivo*[A]					
pET.mp1	*ca*HydE	*ca*HydA	pETDuet-1	Ampicillin	Matthew Posewitz[31]
pCDF.mp2	*ca*HydF	*ca*HydG	pCDFDuet-1	Spectinomycin	Matthew Posewitz[31]
pACYC.ew3	*da*PFOR		pACYCDuet-1	Chloramphenicol	This work
pACYC.ew4	*da*PFOR	*so*FD	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew5	*da*PFOR	*zm*FD	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew6	*da*PFOR	*cr*FD	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew7	*da*PFOR	*ca*FD	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew8	*so*FD		pACYCDuet-1	Chloramphenicol	This work
pACYC.ew9	*zm*FD		pACYCDuet-1	Chloramphenicol	This work
pACYC.ew10	*cr*FD		pACYCDuet-1	Chloramphenicol	This work
pACYC.ew11	*ca*FD		pACYCDuet-1	Chloramphenicol	This work
FNR-supported growth[B]					
pCDF.ew12	*zm*FNR		pCDFDuet-1	Spectinomycin	This work
pACYC.ew13	*so*FD	*zm*SIR	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew14	*zm*FD	*zm*SIR	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew15	*cr*FD	*zm*SIR	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew16	*ca*FD	*zm*SIR	pACYCDuet-1	Chloramphenicol	This work
pACYC.ew17	*zm*SIR		pACYCDuet-1	Chloramphenicol	This work
Hydrogenase-supported growth and selection[C]					
pACYC.ew18	*cr*HydEF	*cr*HydG	pACYCDuet-1	Chloramphenicol	This work
pET.ew19	*so*FD	*zm*SIR	pETDuet-1	Ampicillin	This work
pCDF.ew20	*cr*HydA		pCDFDuet-1	Spectinomycin	This work
pCDF.ew21	*ca*HydA		pCDFDuet-1	Spectinomycin	This work
pCDF.ew22	*cs*HydA		pCDFDuet-1	Spectinomycin	This work

Duet vector backbones (Novagen) were used for all protein expression. Complete vector sequences are provided in additional file 1. A) Plasmids mp1-ew11 were used to generate PFOR-driven hydrogenase activity data for Figure 4. B) Plasmids ew12-ew17 were used to assess knockout strain insulation for Figure 2. They were also used to measure ferredoxin performance in Figure 4. C) Plasmids ew18-ew22 facilitated the hydrogenase O_2 tolerance measurements in Figure 3. They were also used in the hydrogenase-supported growth curves in Figure 5 and for the genetic selection.

Media recipes used in Barstow et al 2011

Selective Media
Selective media was a standard M9 formulation, supplemented with additional glucose, sulfate, ferric iron and a rich mix of amino acids less cysteine and methionine.

15 g	Agar
12.8 g	$Na_2HPO_4 \cdot 7H_2O$
3 g	KH_2PO_4
0.5 g	NaCl
1 g	NH_4Cl
0.1 mL	1M $CaCl_2$
10 mL	1M $MgSO_4$
25 mg	Ferric citrate
0.5 g	Sulfur dropout powder
100 mL	20% Glucose
1M IPTG	1mL

Sulfur dropout powder is a rich supplement mix with cysteine and methionine omitted:

Amino acids			
Alanine	2.0 g	Leucine	10.0 g
Arginine	2.0 g	Lysine	2.0 g
Asparagine	2.0 g	Methionine	0.0 g
Aspartic acid	2.0 g	Phenylalanine	2.0 g
Cysteine	0.0 g	Proline	2.0 g
Glutamic acid	2.0 g	Serine	2.0 g
Glutamine	2.0 g	Threonine	2.0 g
Glycine	2.0 g	Tryptophan	2.0 g
Histidine	2.0 g	Tyrosine	2.0 g
Isoleucine	2.0 g	Valine	2.0 g

Nucleotide bases	
Adenine	0.5 g
Uracil	2.0 g
Vitamins	
p-Aminobenzoic acid	0.2 g
Inositol	2.0 g
Adenine	0.5 g
Uracil	2.0 g

Induction media
Induction media for hydrogenase expression was LB with added glucose, ferric iron, phosphate buffer and Baker's antifoam reagent.

5 g	Bacto tryptone
2.5 g	Bacto yeast extract
5 g	Sodium chloride
1.2 g	KH_2PO_4
7.2 g	K_2HPO_4
2.5 mL	20% Glucose
12.5 mg	Ferric citrate
1 mL	1M IPTG
50 µL	Baker's antifoam B

Appendix B

Harnessing nature's toolbox: regulatory elements for synthetic biology[1]

[1] Originally published as:
Boyle, P. M. & Silver, P. A. Harnessing nature's toolbox: regulatory elements for synthetic biology. *Journal of the Royal Society, Interface / the Royal Society* **6 Suppl 4**, S535–46 (2009). Reprinted with permission from the editor.

Interface focus

The Knowledge Network

FirstCite® e-publishing

J. R. Soc. Interface
doi:10.1098/rsif.2008.0521.focus
Published online

REVIEW

Harnessing nature's toolbox: regulatory elements for synthetic biology

Patrick M. Boyle and Pamela A. Silver*

Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, WAB 536, Boston, MA 02115, USA

Synthetic biologists seek to engineer complex biological systems composed of modular elements. Achieving higher complexity in engineered biological organisms will require manipulating numerous systems of biological regulation: transcription; RNA interactions; protein signalling; and metabolic fluxes, among others. Exploiting the natural modularity at each level of biological regulation will promote the development of standardized tools for designing biological systems.

Keywords: biological regulation; natural modularity; standardized tools

1. INTRODUCTION

By analogy to other engineering disciplines, synthetic biologists aim to construct complex 'devices' as assemblies of well-defined modular parts. The synthetic biology approach is predicated on the idea that modular biological elements exist, which can be repurposed or modified for the construction of new devices (Drubin *et al.* 2007).

The unique capabilities of biological organisms make them an attractive target for engineering. Living cells are self-replicating, self-repairing chemical factories that can, in principle, be reconfigured by altering their DNA blueprint. The development of DNA synthesis technology has allowed researchers to create any DNA sequence they desire, even entire genomes (Gibson *et al.* 2008). The ability to synthesize DNA promises to make the entire repertoire of known biological diversity available to synthetic biologists.

Despite this power, rationally designed biological devices rarely function entirely as predicted, and are often less robust than natural systems. With the price of DNA synthesis falling and worldwide interest in synthetic biology rising, the progress of synthetic biology has not been limited by funding or ambition. Instead, progress made in the last decade of synthetic biology research has revealed the unique difficulties of engineering living systems. In particular, endogenous regulatory systems often interfere with the function of synthetic biological devices (Arkin & Fletcher 2006).

Improving our ability to engineer biology will require an improved ability to predictably regulate biological systems. Natural biological systems are regulated at many levels: transcription; RNA processing; translation; protein–protein interactions; and protein–substrate interactions all exert control on cellular processes. The spatial and temporal organization of these myriad control systems is vital to their proper function. The fact that RNA interference, an integral control mechanism in eukaryotes, was only discovered in the late 1990s demonstrates that we are still discovering fundamental building blocks of biological systems (Fire *et al.* 1998).

The identification of biological modules with regulatory functions will allow the construction of more complex devices. For decades, promoters have been used to control the expression of recombinant genes (Reznikoff *et al.* 1969; Casadaban 1975). Similarly, modular protein elements, such as zinc fingers, allow the design of novel protein interactions (Drubin *et al.* 2007). Harnessing biological modularity provides insight into engineering biological regulation. In this review, we will explore recent advances in the field of synthetic biology, with emphasis on the development of biological modules for the regulation of synthetic devices.

2. TRANSCRIPTIONAL SIGNAL PROCESSING

It has long been understood that transcription factors target specific DNA elements to control gene expression (Jacob & Monod 1961). Promoters are modular DNA elements that can be used to drive the transcription of a gene, a property that has often been exploited for biological research (Casadaban 1975). As a well-understood mechanism of biological regulation, transcriptional control has been a feature of many synthetic devices.

*Author for correspondence (pamela_silver@hms.harvard.edu).

One contribution to a Theme Supplement 'Synthetic biology: history, challenges and prospects'.

Received 9 December 2008
Accepted 4 February 2009

Table 1. Basic two-input logic gates. (Each gate produces a single true or false output based on the state of two inputs (Irving 1961). F, 'false'; T, 'true'; NAND, 'not AND'; NOR, 'not OR'.)

	input		
gate	A	B	output
AND	F	F	F
	T	F	F
	F	T	F
	T	T	T
OR	F	F	F
	T	F	T
	F	T	T
	T	T	T
NAND	F	F	T
	T	F	T
	F	T	T
	T	T	F
NOR	F	F	T
	T	F	F
	F	T	F
	T	T	F

Transcriptional control can be exploited for signal integration, i.e. processing multiple inputs and producing well-defined outputs. The operations that signal integration elements perform can be represented as logic gates. Boolean logic gates integrate two or more input signals and output one of two values (Irving 1961). For example, a two-input AND gate returns an output of 'true' when both inputs are true, and returns 'false' when either or both of the inputs are false (table 1). Transcriptional devices can serve as logic gates by producing outputs based on the state of input promoters; an AND gate would produce a certain output only if both input promoters were induced.

Ideally, signal integration elements and promoter inputs would be functionally separate in a synthetic device. For example, replacing an ara promoter with a lac promoter should modify the specificity of the device from arabinose to lactose without affecting how the rest of the device functions. Towards this end, a 'modular' AND gate was developed, allowing the modification of the promoter inputs and the gate output, leaving the core logic module intact (figure 1*a*; Anderson *et al.* 2007). Two promoters receive input to the device: one drives the expression of a T7 polymerase containing two amber mutations (T7ptag), while the other directs expression of the amber-suppressing tRNA supD. When both promoters are activated, supD expression allows the proper translation of T7ptag, which in turn transcribes a reporter gene with a T7 promoter. The original input promoters were induced by salicylate and arabinose for the expression of supD and T7ptag, respectively, with green fluorescent protein (GFP) as the output reporter.

This supD/T7ptag AND gate was intended to be modular, in that the device functions as an AND gate regardless of the input promoters or the output protein. To confirm modular function, the device inputs were replaced with magnesium-repressed and AI-1-inducible (LuxR quorum sensor) promoters, and the output was replaced with the invasin gene that triggers the invasion of mammalian cells. Although the initial reconfiguration of the device did not perform as an AND gate, tuning expression by varying the ribosome-binding sites of each promoter restored functionality. The new device allows *Escherichia coli* cells to invade mammalian cells in the case that exogenous magnesium is absent and the AI-1 signal is present. Importantly, the signal integration portion of the device was unmodified, and only minimal tuning was required to restore logic gate function after vastly altering the device inputs and outputs.

Producing high-fidelity outputs is as important as proper signal integration. Ideally, a synthetic device will remain in a persistent output state until the input state is changed. Genetic toggle switches have been demonstrated, incorporating transcriptional feedback loops to ensure that the device remains in one of two stable states (figure 1*b*; Gardner *et al.* 2000). Devices that feature oscillatory outputs based on transcriptional feedback have also been constructed (Elowitz & Leibler 2000; Stricker *et al.* 2008). 'Memory devices' feature a persistent output in response to a transient input; they are activated by a specific input and remain active after the input is removed (Ajo-Franklin *et al.* 2007).

A memory device constructed in the yeast *Saccharomyces cerevisiae* transmits a fluorescent response for many generations following a transient stimulus (figure 1*c*; Ajo-Franklin *et al.* 2007). The device features two fluorescently labelled genes: a red fluorescent protein (RFP)-tagged sensor gene, which, when expressed, activates the transcription of a yellow fluorescent protein (YFP)-tagged auto-feedback gene. The auto-feedback gene possesses the same transcriptional activator module as the sensor gene, thus the auto-feedback gene activates its own expression following the transient stimulus. Experimentation and quantitative models of the device suggested that tight regulation of transcription was required to maintain memory. Leaky promoters would trigger activation of the device in the absence of inducer, and the concentration of the auto-feedback protein following induction needs to remain above the threshold for maintaining memory following cell division. The final device exhibits the required bistability, maintaining the YFP signal even after the inducer was removed.

3. RNA SIGNAL PROCESSING

RNA is a versatile molecule with a variety of roles in cellular functions. The ability of RNA to transmit genetic information as well as conduct enzymatic catalysis has led to the hypothesis that early life used RNA in lieu of DNA and proteins (Bartel & Unrau 1999). Regulatory RNAs can exert control by antisense binding with mRNA, by conformational changes that disrupt transcription or translation or by ribozyme activity (Isaacs *et al.* 2006; Saito & Inoue 2008). In the context of synthetic biology, many devices have been constructed that take advantage of RNA-mediated regulation.

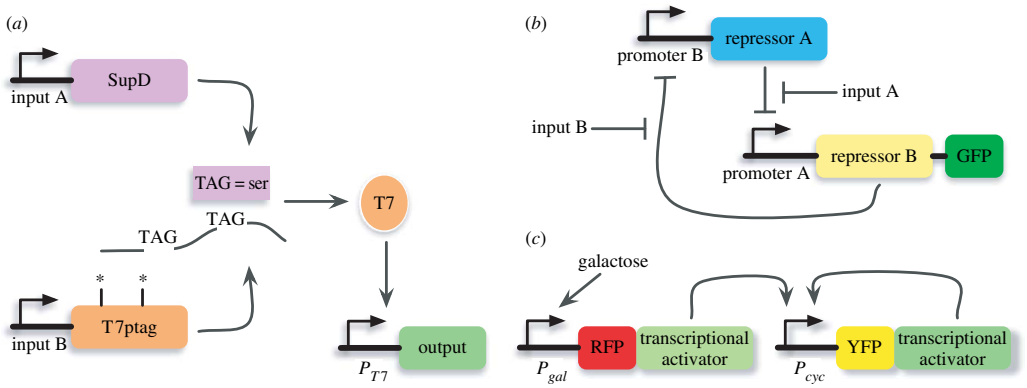


Figure 1. Synthetic devices based on transcriptional logic. (*a*) A modular AND gate. Two input promoters control the expression of the SupD tRNA suppressor and a T7 polymerase with two TAG stop codons, denoted by asterisks. Activation of both input promoters allows SupD to suppress the early stop codons by inserting serine, and the functional T7 protein activates expression of the reporter (Anderson *et al.* 2007). (*b*) A genetic toggle switch. Bistability is achieved via two mutually repressing genes. Addition of inducer molecules allows switching between stable states (Gardner *et al.* 2000). (*c*) A cellular memory device. Galactose induces expression of an RFP-tagged transcriptional activator that triggers expression of a YFP-tagged reporter. The YFP-tagged protein then activates itself, maintaining YFP expression (Ajo-Franklin *et al.* 2007).

Riboswitches are regulatory RNAs that bind small molecules or peptides via a specific aptamer domain (Patel *et al.* 1997). Natural riboswitches can be found in the 5′ untranslated region (UTR) of mRNA, where ligand binding triggers a conformational change that can negatively or positively modulate transcription or translation (Mandal & Breaker 2004; Tucker & Breaker 2005; Isaacs *et al.* 2006). Catalytic RNAs such as hammerhead ribozymes (HHRz; Khvorova *et al.* 2003) can even be attached to riboswitches to construct ligand-dependent ribozymes (Win & Smolke 2008).

In the context of synthetic devices, riboswitches allow promoter-independent control of gene expression. For example, RNA 'antiswitches' have been engineered in *S. cerevisiae* (Bayer & Smolke 2005). An antiswitch is an RNA molecule containing a ligand-binding aptamer domain and an antisense regulator domain. Antiswitches can be designed to activate or repress translation in response to ligand binding. 'Off antiswitches' feature an antisense domain that is stabilized as a stem loop in the absence of ligand, permitting translation of the targeted mRNA. Ligand binding to the aptamer domain triggers a conformational change, exposing the antisense domain and repressing translation. 'On antiswitches', on the other hand, feature an antisense domain that forms a stem loop when the ligand is bound, releasing the mRNA target from repression. Importantly, the aptamer and antisense domains are functionally modular; exchanging a theophylline aptamer domain for a tetracycline aptamer domain changes the specificity of an antiswitch to tetracycline without altering the antisense targeting.

The modularity of riboswitches has been exploited to construct logic gates (Win & Smolke 2008). Riboswitch logic gates incorporate 'sensor', 'transmitter' and 'actuator' domains (figure 2*a*). The sensor domain consists of a ligand-dependent aptamer, and the conformational change triggered by ligand binding alters the conformation of the transmitter domain. Attaching the sensor and transmitter to an actuator domain, in this case an HHRz, confers ligand-dependent ribozyme activity. The devices were embedded in the 3′ UTR of mRNA transcripts, such that HHRz activation triggers mRNA cleavage leading to decreased expression. More complex devices were constructed by linking pairs of sensor domains, linked by transmitters, to a single actuator. Rearrangement of dual sensor domains relative to an actuator domain yielded AND, NOR, NAND and OR gates (figure 2*b*).

Other devices were also constructed by placing two sensor–actuator pairs in series on an mRNA. One such device is a 'bandpass filter', so named because expression of the reporter protein only occurs at intermediate inducer concentrations, with low and high inducer concentrations leading to mRNA degradation. This was accomplished by including a 'buffer gate' that prevents ribozyme cleavage in the presence of theophylline as well as an 'inverter gate' on the same mRNA that activates ribozyme cleavage in the presence of theophylline.

Boolean logic gates have also been designed with small interfering RNAs (siRNAs; Rinaudo *et al.* 2007). In these devices, inputs trigger siRNA expression, with each siRNA targeting a specific mRNA. The targeted mRNA molecules each encode a repressor protein such as LacI that repress the expression of a fluorescent protein reporter. For example, an AND gate consists of two mRNAs driven by the same promoter, but with different 3′ UTR targets, A and B. Each mRNA encodes LacI or LacI–KRAB to repress the expression of the reporter protein. The presence of both siRNA A and siRNA B is required to knock down the mRNAs and allow reporter expression. An OR gate was constructed by expressing a single mRNA for the repressor protein, with two 3′ UTR targets, such that siRNA A or siRNA B is sufficient for knockdown.

This siRNA-based logic allowed the construction of devices that evaluate complex expressions such as

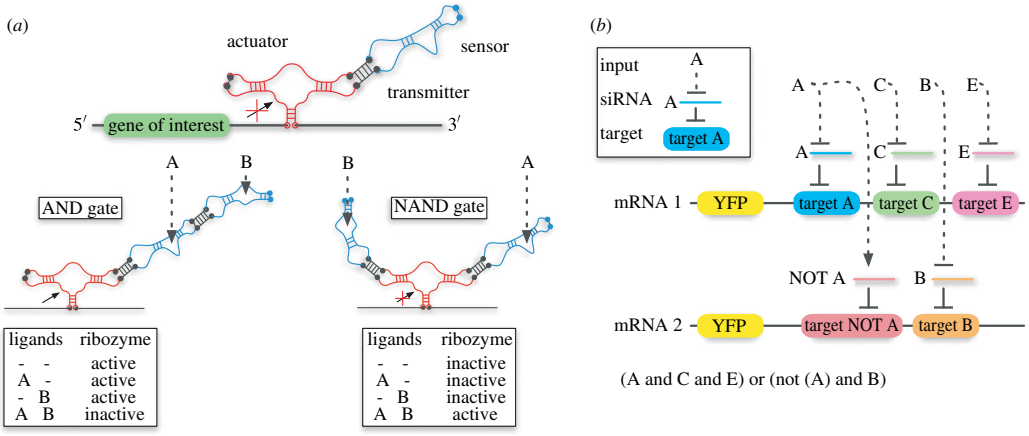


Figure 2. RNA-based logic gates. (*a*) Modular ribozymes consist of 'actuators' (red) that contain the HHRz ribozyme; 'transmitters' that transmit conformational changes to the actuator; and 'sensors', the RNA aptamers that bind ligands. An arrow next to the actuator denotes that the ribozyme is active in the absence of ligand. A crossed out arrow denotes that the ribozyme is inactive in the absence of ligand. Multiple sensor domains can be added to either end of the actuator. An example AND gate and NAND gate are shown, multiple actuators can also be incorporated in series on a single mRNA (Win & Smolke 2008). (*b*) Evaluation of complex logic in an RNAi-based device. Input A represses transcription of the 'A' siRNA and activates transcription of the 'NOT A' siRNA. Inputs C, B and E repress transcription of their respective siRNAs. Each siRNA downregulates the expression of its target mRNA if expressed (Rinaudo *et al.* 2007).

'(A AND B AND C) OR (D AND E)' and '(A AND C AND E) OR (NOT(A) AND B)' (figure 2*b*), with A, B, C, D and E representing different inputs. Every possible input combination was experimentally verified, with the second device yielding the only incorrect output, returning false in the case that A and E were true and B and C were false (Rinaudo *et al.* 2007).

RNA logic devices demonstrate that RNA can regulate complex synthetic networks. Both the riboswitch and RNAi-based devices take advantage of antisense base pairing to confer specificity between synthetic RNA inputs and their mRNA targets. Modular riboswitches may prove to be portable to a wide range of host species, as RNA folding and ribozyme activity occur independently of the host cellular machinery. Complex RNA regulatory devices in natural systems, such as the *S*-adenosyl-methionine/adenosylcobalamin NOR gate in *Bacillus clausii*, demonstrate that RNA aptamers are an effective means of responding to changes in metabolite concentrations (Nahvi *et al.* 2002; Stoddard & Batey 2006).

In higher eukaryotes, introns appear to be an important regulatory element at the mRNA level. Approximately 95 per cent of the average human gene is made up of introns (Lander *et al.* 2001; Venter *et al.* 2001). Both introns and exons must be transcribed; introns are removed from the mRNA prior to translation. In the development of multicellular organisms, one suggested role for introns is in regulating the timing of gene expression, as intron processing increases the time between transcription and translation. Time delay coupled with autoinhibition can produce oscillations, and oscillatory gene expression is often observed in development (Swinburne & Silver 2008).

To study the role of intron length in development, an intron-containing device was constructed (Swinburne *et al.* 2008). The device contained an intron, a fluorescent reporter and the Tet repressor, which negatively inhibited its own expression. The device demonstrated pulses of expression in mammalian cells, and the frequency of the pulses was dependent on intron length. As in the endogenous genes of higher eukaryotes, introns may provide an additional layer of control to synthetic devices.

4. SYNTHETIC PROTEIN SIGNALLING

Proteins do not interact the way components on a circuit board do. Identical components such as resistors can be placed on the same circuit board and not interfere with one another, because the wiring keeps them connected only to components they were intended to interact with. Proteins, however, can diffuse throughout the cellular compartment that contains them, interacting with any suitable binding partners.

Evolution has found a solution to orthogonal signalling that still allows cells to use the same protein components for multiple processes (Bhattacharyya *et al.* 2006). Classes of proteins, such as kinases, share a common mechanism of action but can act on a variety of targets. This is often achieved by varying combinations of adapter domains and effector domains. Signalling protein interactions are often mediated by events such as phosphorylation that change binding affinities and 'rewire' the network (Pawson 2007).

Prokaryotic two-component signalling systems provide a simple model system for studying protein interactions. Canonical two-component systems consist of a membrane-bound receptor with ligand-dependent histidine kinase (HK) activity and a response regulator (RR) protein, usually a transcription factor (West & Stock 2001).

The rational design of two-component signalling systems has recently been demonstrated (Skerker *et al.* 2008). Multiple sequence alignments of HK and RR pairs identified residues that covaried, representing interacting partners. Modification of as few as three interacting residues in the HK EnvZ (an *E. coli* osmolarity sensor) switched the EnvZ HK phosphorylation specificity from its cognate RR OmpR to the non-cognate RR RstA, which is not normally induced by osmotic changes. Additionally, the RR specificity of EnvZ was switched to that of CC1181, a *Caulobacter crecentus* sensor protein. This research establishes a protocol for the rational design of two-component systems, as well as validating methods of HK–RR interaction prediction. The existence of two-component systems in eukaryotes (Saito 2001) implies that novel HK–RR pairs could augment eukaryotic devices as well.

Eukaryotic signalling cascades are usually more complex than two-component systems, but eukaryotic signalling proteins can still be reprogrammed to accept new inputs. For example, a guanine nucleotide exchange factor (GEF) was modified to be responsive to protein kinase A (PKA; Yeh *et al.* 2007). Active GEFs catalyse the exchange of GDP for GTP bound to Rho GTPases, and, in turn, GTP-bound Rho activates downstream effectors (Rossman *et al.* 2005). To build a synthetic GEF, researchers replaced the cognate autoinhibitory domain of the CDC42-specific GEF Itsn1 with a PKA-dependent autoinhibitory domain, leaving the GEF catalytic domain intact (figure 3; Yeh *et al.* 2007). The addition of forskolin, a PKA activator, induces production of filopodia in mammalian cells, indicating Itsn1 signalling. Substituting different GEF catalytic domains also produced new signalling behaviours, and signalling cascades involving two synthetic GEFs were also functional (Yeh *et al.* 2007).

Protein domains can be combined to produce novel switch-like behaviour. A chimeric protein with two separate ligand-binding domains could act as a switch or an OR gate if only one ligand-binding domain could be occupied at a time. To isolate new protein switches, researchers overlapped functional ligand-binding domains and peptides in chimeric proteins, such that correct folding of one domain would disrupt folding of the other (Sallee *et al.* 2007). Out of 25 candidates, seven chimeric proteins yielded functional switches, with domains that are unstructured in the absence of ligand showing the highest likelihood of success (Sallee *et al.* 2007).

Protein ligands that bind cell surface receptors can also be used as modular regulatory elements. One such device, a chimeric-activating protein, was constructed by connecting an epidermal growth factor (EGF) ligand and an interferonα-2a (IFNα-2a) ligand via a flexible linker (Cironi *et al.* 2008). The EGF ligand acts as a targeting element, binding the EGF receptor (EGFR). The IFNα-2a ligand triggers the desired action of the device, binding IFNα-2a–IFNα receptor 2 (IFNAR2) and activating the Jak–Stat pathway (Platanias 2005). In the chimeric activator, the IFNα-2a ligand was mutated to reduce its binding affinity for IFNAR2. Reducing IFNα-2a binding affinity had the desired effect: IFNα-2a-mediated activation of the

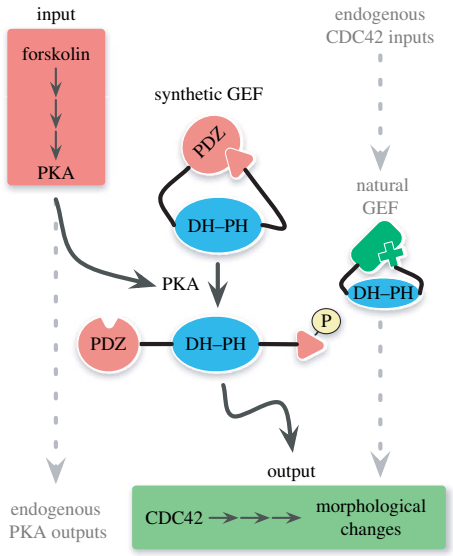


Figure 3. A synthetic GEF re-routes PKA signalling to activate the CDC42 pathway. The Dbl homology–pleckstrin homology domain (DH–PH) from a GEF involved in CDC42 signalling is combined with a PDZ domain and a PKA target peptide. The PDZ domain is normally bound to the target peptide, inactivating the DH–PH. Forskolin activates PKA that phosphorylates the target peptide. Phosphorylation of the target allows the DH–PH to activate CDC42 signalling (Yeh *et al.* 2007).

Jak–Stat pathway occurs only when both EGFR and IFNAR2 were present on the cell surface. EGF binding to EGFR brings IFNα-2a closer to the cell surface, increasing the likelihood of IFNα-2a–IFNAR2 binding and subsequent Jak–Stat signalling. As well as having therapeutic applications, chimeric activators could be incorporated into synthetic devices for intercellular signalling (Cironi *et al.* 2008).

These synthetic protein devices demonstrate that the rearrangement of natural protein modules can yield new behaviours. To achieve protein–protein interactions that are truly orthogonal to an existing protein network, however, it may be necessary to design new protein interactions. There is evidence that new types of signalling, such as tyrosine kinase signalling, evolved in response to the saturation of previous signalling networks (King *et al.* 2003; Bhattacharyya *et al.* 2006). The engineering of novel signalling systems may permit synthetic devices to operate in cells without the interference of the endogenous protein network.

Modification of protein interfaces and binding pockets (as opposed to the rearrangement of modular elements discussed previously) has proven to be an effective method of altering protein specificity. Computational modelling of protein interfaces is often similar to *ab initio* modelling of whole proteins, although the scope of the model is reduced to the interface in question (Kortemme & Baker 2004). By modifying the interacting surface of one protein and predicting

compensating mutations in a binding partner protein, natural protein–protein interfaces have been successfully redesigned (Kortemme *et al.* 2004). Similarly, the rational design of ligand-binding pockets led to engineered periplasmic binding protein receptors capable of binding trinitrotoluene, L-lactate and serotonin (Looger *et al.* 2003).

5. METABOLIC ENGINEERING

The rational design of organisms to produce important metabolites such as biofuels and drugs has been labelled as the defining application of synthetic biology (Brenner *et al.* 2006). Rational reconfiguration of metabolism is also an enormous challenge; metabolite levels are regulated in many ways, and small changes in gene regulation can have amplified effects on the metabolome (Raamsdonk *et al.* 2001; Kell 2006). As demonstrated in the following examples, the construction of a 'metabolic device' requires substantial rewiring of the host cell.

A significant effort in metabolic engineering has been the production of amorphadiene in *E. coli* (Martin *et al.* 2003) and artemisinic acid in *S. cerevisiae* (Ro *et al.* 2006), both precursors to the anti-malarial drug artemisinin. In both cases, the amorphadiene synthase (ADS) gene from *Artemisia annua* L was heterologously expressed, and flux through the host metabolic network was redirected towards ADS. In the case of *S. cerevisiae*, expressing the ADS is sufficient for amorphadiene production, albeit with low yields. Adjusting the expression levels of five genes involved in the production of farnesyl pyrophosphate (FPP, converted to amorphadiene by ADS) yielded a 500-fold increase in amorphadiene production (figure 4). Screening a library of *A. annua* cytochrome P450 expressed sequence tags yielded an enzyme that catalysed the conversion of amorphadiene to artemisinic acid, which was then integrated into the engineered strain along with NADPH : cytochrome P450 oxidoreductase. In the case of artemisinic acid, as well as in many other metabolic engineering efforts, the redirection of flux via adjustments in gene regulation was paramount in achieving commercially viable yields (Keasling 2008).

Re-routing metabolic flux via gene deletions can force cells to produce more of a desired product. Genome-scale simulations of cellular metabolism such as flux balance analysis (FBA), which we will explore in §6, can predict beneficial gene deletions. This approach has been validated for several metabolic engineering efforts, including the production of lycopene in *E. coli* (Alper *et al.* 2005), and in the case of our own efforts to produce formic acid in *S. cerevisiae* (Kennedy *et al.* in preparation). Results in both studies suggested that further regulatory modifications would boost yields. *In silico* strain design coupled with tight regulation of enzyme levels by synthetic devices will undoubtedly be essential to future metabolic engineering efforts.

As in protein signalling networks, modular manipulation of metabolic enzymes would facilitate the development of new pathways. However, unlike signalling proteins, metabolic enzymes are rarely structurally

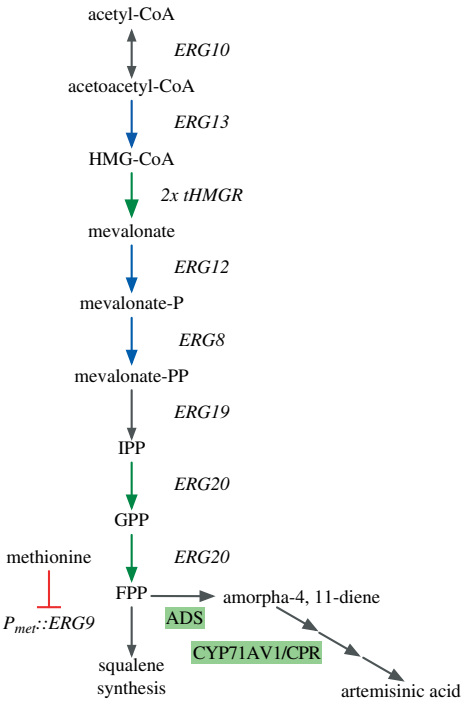


Figure 4. Engineered pathway for artemisinic acid production in *S. cerevisiae*. Blue arrows indicate enzymes indirectly upregulated by expression of *upc2-1*. Green arrows indicate enzymes that were directly upregulated. The red repression arrow indicates that *ERG9* was placed under the control of a methionine-repressed promoter, reducing flux to squalene synthesis. Green boxes indicate the exogenous enzymes ADS and CYP71AV1 and the redox partner protein CPR (Ro *et al.* 2006). CoA, coenzyme A; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; IPP, isopentenylpyrophosphate; GPP, geranyl pyrophosphate.

modular. For example, metabolic enzymes often contain allosteric regulatory sites and the active site within a single domain, confounding efforts to decouple allosteric regulation and catalysis (Bhattacharyya *et al.* 2006). A notable exception is enzymes such as polyketide synthases (PKS), which are not only modular, but their spatial arrangement can also be exploited to create new products.

Rational engineering of PKS assembly lines would open up new possibilities for the synthesis of organic molecules. Polyketides are assembled by linear complexes of PKS proteins, with each PKS performing catalysis on the growing polyketide chain (figure 5). Evolutionary rearrangement of PKS modules has generated a diverse array of natural products, including many antibiotics (Robinson 1991). PKS proteins possess N- and C-terminal 'docking domains' for attachment to other PKS (Thattai *et al.* 2007). Combinatorial shuffling of PKS modules resulted in the *in vivo* synthesis of novel polyketides (Menzella *et al.* 2005). Computational modelling of potential PKS products

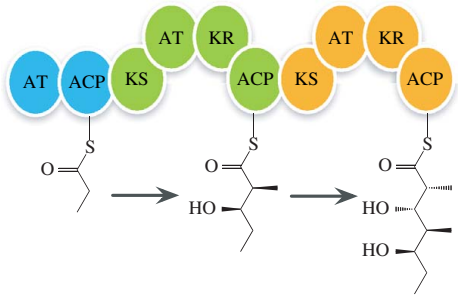


Figure 5. The DEBS1 PKS that catalyses the first steps in synthesizing 6-deoxyerythronolide B. Blue domains comprise the loading module, containing an acyl transferase (AT) and an acyl carrier protein (ACP). Two extender modules, in green and orange, each contain a ketosynthase (KS), an AT, a ketoreductase (KR) and an ACP (Menzella *et al.* 2005).

suggests that billions of possible molecules could be synthesized via engineered PKS combinations, and that it may be possible to predict PKS combinations that will produce a desired compound (González-Lergier *et al.* 2005).

Our ability to conduct 'retro-biosynthesis', the rational design of biological routes to target compounds, is limited both by our knowledge of enzyme properties and our ability to override regulation in biological systems (Prather & Martin 2008). The diversity of compounds synthesized by natural organisms suggests that biological chassis are an ideal platform for chemical production. It has been noted that the development of synthetic biology is paralagous to the development of synthetic chemistry; the synthesis of many important organic compounds was achieved before chemists understood covalent bonds (Yeh & Lim 2007). Similarly, efforts to re-engineer metabolism will contribute to a more complete understanding of metabolic systems and cellular regulation.

6. METABOLIC MODELLING

The state of the cellular metabolic network is a function of the network topology, the physical properties of enzymes and the regulation of enzyme levels and activity. Rational design of metabolism will require accurate models of the metabolic network and how the network is regulated. The complexity of metabolism has necessitated trade-offs in the formulation of metabolic models. In general, current models of metabolism fall into one of two categories: constraint-based models and kinetic models.

Most constraint-based models of metabolism are based on the framework of FBA, a technique that simulates the entire metabolic network of an organism (Varma & Palsson 1994). The only required parameter for an FBA model is a stoichiometric matrix that contains all known metabolic reactions of an organism. Constraints are placed on certain fluxes, defining nutrient availability and relative uptake rates as well as thermodynamic constraints on the reversibility of reactions. It is assumed that, at steady state, the net flux of the system is fixed. The model is then solved for the optimization of an objective function such as maximization of biomass. Since FBA models do not consider enzymatic parameters beyond the stoichiometry of each reaction, the availability of comprehensive databases such as the Kyoto Encyclopedia of Genes and Genomes (http://www.kegg.com) has fostered the development of FBA models for many organisms (Varma & Palsson 1994; Duarte *et al.* 2004; Becker & Palsson 2005; Feist *et al.* 2007; Lee *et al.* 2008; Senger & Papoutsakis 2008). Owing to the genome-scale nature of constraint-based models, *in silico* screens have been applied to predictions of gene essentiality (Edwards & Palsson 2000; Thiele *et al.* 2005; Samal *et al.* 2006; Becker & Palsson 2008) and the related metabolic engineering problem of predicting gene knockouts for strain optimization (Burgard *et al.* 2003; Alper *et al.* 2005; Kennedy *et al.* in preparation).

The successful application of FBA in a variety of organisms demonstrates the use of constraint-based models in the context of metabolism. However, even in the unlikely case that enzyme kinetics are unimportant to determine metabolic flux, traditional FBA models assume that a cell's entire complement of enzymes is available at all times. Regulation can be modelled implicitly, via methods such as minimization of metabolic adjustment, which assumes that regulation will force mutant flux distributions to be as similar to the wild-type distribution as possible (Segrè *et al.* 2002). Models such as regulatory FBA attempt to explicitly model regulation by switching fluxes on and off, based on the experimental data of enzyme expression in various growth conditions (Covert *et al.* 2001; Covert & Palsson 2002; Herrgård *et al.* 2006).

Experimental efforts to modify transcription factor behaviour have underscored the importance of regulation to metabolic fluxes. One such approach, known as 'global transcription machinery engineering', yields improved strains by screening transcription factor mutants. For example, a more ethanol-tolerant strain of *S. cerevisiae* was isolated by mutagenizing the TATA-binding protein *SPT15* (Alper *et al.* 2006). The global influence of transcription factors makes them a powerful tool for strain construction; unfortunately, this has also made it difficult to predict the full effects of transcription factor modification. Future genome-scale models may include genome-scale models of transcriptional regulation, but comprehensive information on transcription factor interactions is not presently available.

Metabolic regulation *in vivo* is influenced by enzyme and substrate concentrations and the kinetic parameters of each enzyme. Kinetic models sacrifice the genome scale of constraint-based models in favour of detailed quantitative modelling of specific pathways. Unlike FBA, metabolic control analysis (MCA) accounts for the kinetic parameters of all enzymes in the pathway, along with the concentrations of the enzymes and the metabolites involved (Fell 1997). Control coefficients for each enzyme, based on experimentally measured parameters, define the amount of influence an enzyme has on a pathway. A key concept of MCA is multisite modulation,

i.e. there is no 'master' enzyme or rate-limiting step in a pathway; instead, each enzyme in the pathway has a non-zero control coefficient, and the control coefficients of all the enzymes in the pathway sum to 1. Thus, MCA can identify enzymes with high control coefficients, and quantitatively predict the impact of adjusting the concentration or rate of those enzymes. Unfortunately, the detailed nature of MCA models makes them impossible to apply on the genome scale until more comprehensive data on enzyme kinetics have been collected.

Even if 'omics' technologies generate a complete list of parameters for genome-scale MCA, the resulting dataset would still be difficult to model (Schuster 1999). One approach to reducing the complexity of a genome-scale model is to group metabolic fluxes into modules, provided that the groupings are functionally relevant. Although individual metabolic enzymes are rarely composed of modular elements, considering groups of enzymes as higher order modules may yield adequate models for understanding metabolic network kinetics at a genome scale.

There is evidence that higher order modularity exists in metabolic networks. Metabolic networks appear to be both scale free and hierarchical (Ravasz *et al.* 2002). Modularity is inherent in such networks, which implies that metabolic networks contain modules of small reaction networks linked to highly connected hubs (Jeong *et al.* 2000; Guimerà & Nunes Amaral 2005). Modelling approaches such as elementary flux modes analysis (Schuster *et al.* 1999, 2000; Klamt & Stelling 2002) and modular MCA (Schuster *et al.* 1993; Acerenza & Ortega 2007; Poolman *et al.* 2007) attempt to identify modular reaction sets, generating simplified network models. Identifying modular elements of metabolism may reveal generalized methods for manipulating metabolic regulation in engineered organisms.

7. SUBCELLULAR ENGINEERING

In computer science, abstraction allows the construction of higher orders of modularity (Abelson *et al.* 1996). Computer software usually takes inputs and returns outputs without involving the user in the intervening calculations. Within complex computer programs, there is further abstraction between modules, separating elements such as memory management and input/output control. Similarly, subcellular compartments are an evolutionary form of abstraction; eukaryotic cells possess membrane-bound organelles with specialized and well-defined tasks. Regulated transport limits interaction between the organelle and the rest of the cell to a set of defined inputs and outputs. Metabolite and protein interactions within the organelle are abstracted from the cytosolic environment. Although prokaryotes lack membrane-bound organelles, protein-bound compartments such as carboxysomes exist in certain species (Price *et al.* 2008). In the context of synthetic devices, abstraction afforded by subcellular compartments could limit the interference of endogenous regulation with the engineered regulation of the device.

Lysosomes, peroxisomes and many other organelles contain enzymes or metabolites that are harmful to the host cell (Page *et al.* 1998; Yeldandi *et al.* 2000). Metabolic reactions that would be thermodynamically unfavourable in the cytosol are often found in organelles (Feldman & Sigman 1983). Organelles could conceivably harbour engineered pathways that would be incompatible with the cytoplasm.

The specialized machinery of organelles could augment synthetic devices. Chloroplasts and mitochondria, being of endosymbiotic origin (Embley & Martin 2006), are the most sophisticated eukaryotic organelles and have useful properties for engineered systems. These organelles possess subcompartments of their own, the thylakoids in the case of chloroplast and the matrix in the case of the mitochondrion (Frey & Mannella 2000; Mustárdy *et al.* 2008). In addition, both organelles can generate electrochemical gradients between their subcompartments (Dimroth *et al.* 2000).

Although no completed examples of synthetic organelle devices exist, there is reason to believe that engineering organelles is feasible. In the case of many organelles, such as chloroplasts (Soll & Schleiff 2004), mitochondria (Truscott *et al.* 2003) and peroxisomes (Léon *et al.* 2006), the targeting of arbitrary proteins to these compartments has been described. Chloroplasts and mitochondria also have self-contained genomes, although gene integration into their genomes is more complex than nuclear integration (Bonnefoy & Fox 2007; Verma & Daniell 2007).

Genome-wide metabolic models often have difficulties modelling organelle metabolism (Satish Kumar *et al.* 2007). However, since organelles compartmentalize their metabolic reactions, they can be abstracted from the rest of cellular metabolism. Inputs and outputs from a detailed kinetic model of organelle metabolism could be interfaced with a genome-scale constraint-based metabolic model. In this manner, detailed models of the organelle in question could be combined with global models for the rest of the cell.

8. MULTISPECIES DEVICES

In nature, environmental niches are often colonized by microbial consortia rather than a single dominating species. This may confer a group advantage to cooperating species. In the bovine rumen, multispecies microbial biofilms manage to degrade cellulosic biomass (McAllister *et al.* 1994). Engineered co-cultures may be able to achieve similarly difficult tasks (Brenner *et al.* 2008).

Co-cultures have been applied to the breakdown of lignocellulose, an important carbon source for biofuels (Eiteman *et al.* 2008). In current industrial processes, hydrolysis of lignocellulose yields a mix of five-carbon sugars, such as xylose, and six carbon sugars, such as glucose. In the presence of both xylose and glucose, *E. coli* will preferentially feed on glucose first. A co-culture of two *E. coli* strains, one that is deficient in xylose usage and the other that is deficient in glucose usage, breaks down the sugar mixture more effectively than a monoculture (Eiteman *et al.* 2008).

Synthetic biology offers a range of devices that may allow the coordination of gene regulation between two species, such as quorum-sensing cell–cell signalling devices (Weber *et al.* 2007; Balagaddé *et al.* 2008; Brenner *et al.* 2008). The major task of synthetic devices in co-culture systems may be to exert a selective pressure to maintain a co-culture. Advances in metabolic modelling will improve our ability to design dependencies between strains, such as a requirement for cross-feeding.

9. CONCLUSIONS

The defining question of synthetic biology research moving forward will not be whether biology can be engineered, but how to develop engineering principles for biological systems. Understanding natural regulatory systems, developing improved regulatory systems for synthetic devices and properly interfacing synthetic devices with host cells will play a large role in this process.

The synthetic devices presented here have demonstrated that functioning devices can be constructed, even though our understanding of biological systems is incomplete. In the most promising cases, engineering progress has also provided new biological insights. There is also much to be learned from building synthetic devices that do not work as planned. Designing synthetic devices in an iterative fashion, with experimental results allowing improved models and vice versa, will allow increasingly complex device designs. In addition, modelling shortfalls and unexpected experimental outcomes may shed light on new mechanisms of endogenous biological regulation.

The success of synthetic biology endeavours depends heavily on understanding how biological systems are regulated. In natural systems, there are regulated interactions between DNA, RNA, proteins and metabolites. Identifying modular regulatory elements such as promoters and riboswitches has been essential to the progress of synthetic biology. There is considerable evidence that genomic rearrangements and horizontal gene transfer have driven the evolution of new biological capabilities. Similarly, the identification of biological modules that confer new functionality when assembled in different contexts will drive the progress of synthetic biology.

We thank Christina Agapakis and Caleb Kennedy for reviewing the paper. This work was supported by Harvard University Center for the Environment (HUCE) and National Institutes of Health (NIH) funding to P.A.S., and a HUCE Graduate Fellowship and an NIH Cell and Developmental Biology training grant to P.M.B.

REFERENCES

Abelson, H., Sussman, G. J. & Sussman, J. 1996 *Structure and interpretation of computer programs*. Cambridge, MA: The MIT Press.

Acerenza, L. & Ortega, F. 2007 Modular metabolic control analysis of large responses. *FEBS J.* **274**, 188–201. (doi:10.1111/j.1742-4658.2006.05575.x)

Ajo-Franklin, C. M., Drubin, D. A., Eskin, J. A., Gee, E. P. S., Landgraf, D., Phillips, I. & Silver, P. A. 2007 Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276. (doi:10.1101/gad.1586107)

Alper, H., Jin, Y.-S., Moxley, J. F. & Stephanopoulos, G. 2005 Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab. Eng.* **7**, 155–164. (doi:10.1016/j.ymben.2004.12.003)

Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. 2006 Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568. (doi:10.1126/science.1131969)

Anderson, J. C., Voigt, C. & Arkin, A. P. 2007 Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3**, 133. (doi:10.1038/msb4100173)

Arkin, A. P. & Fletcher, D. A. 2006 Fast, cheap and somewhat in control. *Genome Biol.* **7**, 114. (doi:10.1186/gb-2006-7-8-114)

Balagaddé, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., Quake, S. R. & You, L. 2008 A synthetic *Escherichia coli* predator–prey ecosystem. *Mol. Syst. Biol.* **4**, 187. (doi:10.1038/msb.2008.24)

Bartel, D. P. & Unrau, P. J. 1999 Constructing an RNA world. *Trends Cell Biol.* **9**, M9–M13. (doi:10.1016/S0962-8924(99)01669-4)

Bayer, T. S. & Smolke, C. D. 2005 Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.* **23**, 337–343. (doi:10.1038/nbt1069)

Becker, S. A. & Palsson, B. Ø. 2005 Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* **5**, 8. (doi:10.1186/1471-2180-5-8)

Becker, S. A. & Palsson, B. Ø. 2008 Three factors underlying incorrect *in silico* predictions of essential metabolic genes. *BMC Syst. Biol.* **2**, 14. (doi:10.1186/1752-0509-2-14)

Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. A. 2006 Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680. (doi:10.1146/annurev.biochem.75.103004.142710)

Bonnefoy, N. & Fox, T. D. 2007 Directed alteration of *Saccharomyces cerevisiae* mitochondrial DNA by biolistic transformation and homologous recombination. *Methods Mol. Biol.* **372**, 153–166. (doi:10.1007/978-1-59745-365-3_11)

Brenner, K., You, L. & Arnold, F. H. 2008 Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26**, 483–489. (doi:10.1016/j.tibtech.2008.05.004)

Brenner, M. P. *et al.* 2006 Engineering microorganisms for energy production. Report JSR-05-300, US Department of Energy, Washington, DC.

Burgard, A., Pharkya, P. & Maranas, C. 2003 Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84**, 647–657. (doi:10.1002/bit.10803)

Casadaban, M. J. 1975 Fusion of the *Escherichia coli* lac genes to the ara promoter: a general technique using bacteriophage Mu-1 insertions. *Proc. Natl Acad. Sci. USA* **72**, 809–813. (doi:10.1073/pnas.72.3.809)

Cironi, P., Swinburne, I. A. & Silver, P. A. 2008 Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. *J. Biol. Chem.* **283**, 8469–8476. (doi:10.1074/jbc.M708502200)

Covert, M. W. & Palsson, B. Ø. 2002 Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277**, 28 058–28 064. (doi:10.1074/jbc.M201691200)

Covert, M. W., Schilling, C. H. & Palsson, B. Ø. 2001 Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73–88. (doi:10.1006/jtbi.2001.2405)

Dimroth, P., Kaim, G. & Matthey, U. 2000 Crucial role of the membrane potential for ATP synthesis by F(1)F(o) ATP synthases. *J. Exp. Biol.* **203**, 51–59.

Drubin, D. A., Way, J. C. & Silver, P. A. 2007 Designing biological systems. *Genes Dev.* **21**, 242–254. (doi:10.1101/gad.1507207)

Duarte, N. C., Herrgård, M. J. & Palsson, B. Ø. 2004 Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**, 1298–1309. (doi:10.1101/gr.2250904)

Edwards, J. S. & Palsson, B. Ø. 2000 Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinform.* **1**, 1. (doi:10.1186/1471-2105-1-1)

Eiteman, M. A., Lee, S. A. & Altman, E. 2008 A co-fermentation strategy to consume sugar mixtures effectively. *J. Biol. Eng.* **2**, 3. (doi:10.1186/1754-1611-2-3)

Elowitz, M. B. & Leibler, S. 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338. (doi:10.1038/35002125)

Embley, T. M. & Martin, W. 2006 Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630. (doi:10.1038/nature04546)

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V. & Palsson, B. Ø. 2007 A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121. (doi:10.1038/msb4100155)

Feldman, R. I. & Sigman, D. S. 1983 The synthesis of ATP by the membrane-bound ATP synthase complex from medium 32Pi under completely uncoupled conditions. *J. Biol. Chem.* **258**, 12 178–12 183.

Fell, D. A. 1997 *Understanding control of metabolism.* London, UK: Portland Press.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. 1998 Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811. (doi:10.1038/35888)

Frey, T. G. & Mannella, C. A. 2000 The internal structure of mitochondria. *Trends Biochem. Sci.* **25**, 319–324. (doi:10.1016/S0968-0004(00)01609-1)

Gardner, T. S., Cantor, C. R. & Collins, J. J. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342. (doi:10.1038/35002131)

Gibson, D. G. *et al.* 2008 Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220. (doi:10.1126/science.1151721)

González-Lergier, J., Broadbelt, L. J. & Hatzimanikatis, V. 2005 Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J. Am. Chem. Soc.* **127**, 9930–9938. (doi:10.1021/ja051586y)

Guimerà, R. & Nunes Amaral, L. A. 2005 Functional cartography of complex metabolic networks. *Nature* **433**, 895–900. (doi:10.1038/nature03288)

Herrgård, M. J., Lee, B. S., Portnoy, V. & Palsson, B. Ø. 2006 Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* **16**, 627–635. (doi:10.1101/gr.4083206)

Irving, A. 1961 *Thinking machines, a layman's introduction to logic, Boolean algebra, and computers.* New York, NY: John Day Co.

Isaacs, F. J., Dwyer, D. J. & Collins, J. J. 2006 RNA synthetic biology. *Nat. Biotechnol.* **24**, 545–554. (doi:10.1038/nbt1208)

Jacob, F. & Monod, J. 1961 Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. 2000 The large-scale organization of metabolic networks. *Nature* **407**, 651–654. (doi:10.1038/35036627)

Keasling, J. D. 2008 Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* **3**, 64–76. (doi:10.1021/cb7002434)

Kell, D. B. 2006 Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov. Today* **11**, 1085–1092. (doi:10.1016/j.drudis.2006.10.004)

Kennedy, C. J., Boyle, P. M., Waks, Z. & Silver, P. A. In preparation. Systems-level engineering of non-fermentative metabolism in yeast.

Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. 2003 Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10**, 708–712. (doi:10.1038/nsb959)

King, N., Hittinger, C. T. & Carroll, S. B. 2003 Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301**, 361–363. (doi:10.1126/science.1083853)

Klamt, S. & Stelling, J. 2002 Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **29**, 233–236. (doi:10.1023/A:1020390132244)

Kortemme, T. & Baker, D. 2004 Computational design of protein–protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97. (doi:10.1016/j.cbpa.2003.12.008)

Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. 2004 Computational redesign of protein–protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379. (doi:10.1038/nsmb749)

Lander, E. S. *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. (doi:10.1038/35057062)

Lee, J., Yun, H., Feist, A. M., Palsson, B. Ø. & Lee, S. Y. 2008 Genome-scale reconstruction and *in silico* analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Appl. Microbiol. Biotechnol.* **80**, 849–862. (doi:10.1007/s00253-008-1654-4)

Léon, S., Goodman, J. M. & Subramani, S. 2006 Uniqueness of the mechanism of protein import into the peroxisome matrix: transport of folded, co-factor-bound and oligomeric proteins by shuttling receptors. *Biochim. Biophys. Acta* **1763**, 1552–1564. (doi:10.1016/j.bbamcr.2006.08.037)

Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. 2003 Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190. (doi:10.1038/nature01556)

Mandal, M. & Breaker, R. R. 2004 Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463. (doi:10.1038/nrm1403)

Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. 2003 Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802. (doi:10.1038/nbt833)

McAllister, T. A., Bae, H. D., Jones, G. A. & Cheng, K. J. 1994 Microbial attachment and feed digestion in the rumen. *J. Anim. Sci.* **72**, 3004–3018.

Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A. & Santi, D. V. 2005 Combinatorial polyketide biosynthesis by

de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171–1176. (doi:10.1038/nbt1128)

Mustárdy, L., Buttle, K., Steinbach, G. & Garab, G. 2008 The three-dimensional network of the thylakoid membranes in plants: quasihelical model of the granum-stroma assembly. *Plant Cell* **20**, 2552–2557. (doi:10.1105/tpc.108.059147)

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L. & Breaker, R. R. 2002 Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043. (doi:10.1016/S1074-5521(02)00224-7)

Page, L. J., Darmon, A. J., Uellner, R. & Griffiths, G. M. 1998 L is for lytic granules: lysosomes that kill. *Biochim. Biophys. Acta* **1401**, 146–156. (doi:10.1016/S0167-4889(97)00138-9)

Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A. & Nonin, S. 1997 Structure, recognition and adaptive binding in RNA aptamer complexes. *J. Mol. Biol.* **272**, 645–664. (doi:10.1006/jmbi.1997.1281)

Pawson, T. 2007 Dynamic control of signaling by modular adaptor proteins. *Curr. Opin. Cell Biol.* **19**, 112–116. (doi:10.1016/j.ceb.2007.02.013)

Platanias, L. C. 2005 Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386. (doi:10.1038/nri1604)

Poolman, M. G., Sebu, C., Pidcock, M. K. & Fell, D. A. 2007 Modular decomposition of metabolic systems via null-space analysis. *J. Theor. Biol.* **249**, 691–705. (doi:10.1016/j.jtbi.2007.08.005)

Prather, K. L. & Martin, C. H. 2008 *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.* **19**, 468–474. (doi:10.1016/j.copbio.2008.07.009)

Price, G. D., Badger, M. R., Woodger, F. J. & Long, B. M. 2008 Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* **59**, 1441–1461. (doi:10.1093/jxb/erm112)

Raamsdonk, L. M. *et al.* 2001 A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* **19**, 45–50. (doi:10.1038/83496)

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A. L. 2002 Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555. (doi:10.1126/science.1073374)

Reznikoff, W. S., Miller, J. H., Scaife, J. G. & Beckwith, J. R. 1969 A mechanism for repressor action. *J. Mol. Biol.* **43**, 201–213. (doi:10.1016/0022-2836(69)90089-8)

Rinaudo, K., Bleris, L., Maddamsetti, R., Subramanian, S., Weiss, R. & Benenson, Y. 2007 A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25**, 795–801. (doi:10.1038/nbt1307)

Ro, D. K. *et al.* 2006 Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943. (doi:10.1038/nature04640)

Robinson, J. A. 1991 Polyketide synthase complexes: their structure and function in antibiotic biosynthesis. *Phil. Trans. R. Soc. B* **332**, 107–114. (doi:10.1098/rstb.1991.0038)

Rossman, K. L., Der, C. J. & Sondek, J. 2005 GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. *Nat. Rev. Mol. Cell Biol.* **6**, 167–180. (doi:10.1038/nrm1587)

Saito, H. 2001 Histidine phosphorylation and two-component signaling in eukaryotic cells. *Chem. Rev.* **101**, 2497–2509. (doi:10.1021/cr000243+)

Saito, H. & Inoue, T. 2008 Synthetic biology with RNA motifs. *Int. J. Biochem. Cell Biol.* **41**, 398–404. (doi:10.1016/j.biocel.2008.08.017)

Sallee, N. A., Yeh, B. J. & Lim, W. A. 2007 Engineering modular protein interaction switches by sequence overlap. *J. Am. Chem. Soc.* **129**, 4606–4611. (doi:10.1021/ja0672728)

Samal, A., Singh, S., Giri, V., Krishna, S., Raghuram, N. & Jain, S. 2006 Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinform.* **7**, 118. (doi:10.1186/1471-2105-7-118)

Satish Kumar, V., Dasika, M. S. & Maranas, C. D. 2007 Optimization based automated curation of metabolic reconstructions. *BMC Bioinform.* **8**, 212. (doi:10.1186/1471-2105-8-212)

Schuster, S. 1999 Use and limitations of modular metabolic control analysis in medicine and biotechnology. *Metab. Eng.* **1**, 232–242. (doi:10.1006/mben.1999.0119)

Schuster, S., Kahn, D. & Westerhoff, H. V. 1993 Modular analysis of the control of complex metabolic pathways. *Biophys. Chem.* **48**, 1–17. (doi:10.1016/0301-4622(93)80037-J)

Schuster, S., Dandekar, T. & Fell, D. A. 1999 Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* **17**, 53–60. (doi:10.1016/S0167-7799(98)01290-6)

Schuster, S., Fell, D. A. & Dandekar, T. 2000 A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326–332. (doi:10.1038/73786)

Segrè, D., Vitkup, D. & Church, G. M. 2002 Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15 112–15 117. (doi:10.1073/pnas.232349399)

Senger, R. S. & Papoutsakis, E. T. 2008 Genome-scale model for *Clostridium acetobutylicum*: part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* **101**, 1036–1052. (doi:10.1002/bit.22010)

Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M. & Laub, M. T. 2008 Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054. (doi:10.1016/j.cell.2008.04.040)

Soll, J. & Schleiff, E. 2004 Protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* **5**, 198–208. (doi:10.1038/nrm1333)

Stoddard, C. D. & Batey, R. T. 2006 Mix-and-match riboswitches. *ACS Chem. Biol.* **1**, 751–754. (doi:10.1021/cb600458w)

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S. & Hasty, J. 2008 A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519. (doi:10.1038/nature07389)

Swinburne, I. A. & Silver, P. A. 2008 Intron delays and transcriptional timing during development. *Dev. Cell* **14**, 324–330. (doi:10.1016/j.devcel.2008.02.002)

Swinburne, I. A., Miguez, D. G., Landgraf, D. & Silver, P. A. 2008 Intron length increases oscillatory periods of gene expression in animal cells. *Genes Dev.* **22**, 2342–2346. (doi:10.1101/gad.1696108)

Thattai, M., Burak, Y. & Shraiman, B. I. 2007 The origins of specificity in polyketide synthase protein interactions. *PLoS Comput. Biol.* **3**, 1827–1835. (doi:10.1371/journal.pcbi.0030186)

Thiele, I., Vo, T. D., Price, N. D. & Palsson, B. Ø. 2005 Expanded metabolic reconstruction of *Helicobacter*

pylori (iIT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.* **187**, 5818–5830. (doi:10.1128/JB.187.16.5818-5830.2005)

Truscott, K. N., Brandner, K. & Pfanner, N. 2003 Mechanisms of protein import into mitochondria. *Curr. Biol.* **13**, R326–R337. (doi:10.1016/S0960-9822(03)00239-2)

Tucker, B. J. & Breaker, R. R. 2005 Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **15**, 342–348. (doi:10.1016/j.sbi.2005.05.003)

Varma, A. & Palsson, B. Ø. 1994 Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **60**, 3724–3731.

Venter, J. C. *et al.* 2001 The sequence of the human genome. *Science* **291**, 1304–1351. (doi:10.1126/science.1058040)

Verma, D. & Daniell, H. 2007 Chloroplast vector systems for biotechnology applications. *Plant Physiol.* **145**, 1129–1143. (doi:10.1104/pp.107.106690)

Weber, W., Daoud-El Baba, M. & Fussenegger, M. 2007 Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl Acad. Sci. USA* **104**, 10 435–10 440. (doi:10.1073/pnas.0701382104)

West, A. H. & Stock, A. M. 2001 Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem. Sci.* **26**, 369–376. (doi:10.1016/S0968-0004(01)01852-7)

Win, M. N. & Smolke, C. D. 2008 Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460. (doi:10.1126/science.1160311)

Yeh, B. J. & Lim, W. A. 2007 Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **3**, 521–525. (doi:10.1038/nchembio0907-521)

Yeh, B. J., Rutigliano, R. J., Deb, A., Bar-Sagi, D. & Lim, W. A. 2007 Rewiring cellular morphology pathways with synthetic guanine nucleotide exchange factors. *Nature* **447**, 596–600. (doi:10.1038/nature05851)

Yeldandi, A. V., Rao, M. S. & Reddy, J. K. 2000 Hydrogen peroxide generation in peroxisome proliferator-induced oncogenesis. *Mutat. Res.* **448**, 159–177. (doi:10.1016/S0027-5107(99)00234-1)

Appendix C

Parts plus pipes: Synthetic biology approaches to metabolic engineering[1]

[1] Originally published as:
Boyle, P. M. & Silver, P. A. Parts plus pipes: Synthetic biology approaches to metabolic engineering. *Metab Eng* (2011).doi:10.1016/j.ymben.2011.10.003. Reprinted with permission from the editor.

Metabolic Engineering ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Parts plus pipes: Synthetic biology approaches to metabolic engineering

Patrick M. Boyle [a], Pamela A. Silver [a,b,*]

[a] Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[b] Wyss Institute of Biologically Inspired Engineering, Boston, MA 02115, USA

ARTICLE INFO

Keywords:
Synthetic biology
Metabolic engineering
Systems biology
Biological design
Biological engineering

ABSTRACT

Synthetic biologists combine modular biological "parts" to create higher-order devices. Metabolic engineers construct biological "pipes" by optimizing the microbial conversion of basic substrates to desired compounds. Many scientists work at the intersection of these two philosophies, employing synthetic devices to enhance metabolic engineering efforts. These integrated approaches promise to do more than simply improve product yields; they can expand the array of products that are tractable to produce biologically. In this review, we explore the application of synthetic biology techniques to next-generation metabolic engineering challenges, as well as the emerging engineering principles for biological design.

1. Introduction

Engineering cellular metabolism requires an understanding of the metabolic reactions involved as well as the regulatory elements that affect metabolic throughput. Our ability as engineers to modulate metabolic pathways has been augmented in recent years by the influx of methods and biological devices from the field of synthetic biology (Fig. 1).

A primary goal of synthetic biology is to develop engineering principles for biology—to translate a quantitative understanding of biological systems into a methodology for building living devices out of standardized biological parts. The advent of cost effective DNA sequencing and de-novo synthesis has resulted in a tremendous increase in the number of potential biological parts available to synthetic biologists (Boyle and Silver, 2009). The development of assembly standards and open databases has facilitated the development and sharing of these parts (Anderson et al., 2010; Knight, 2003; Phillips and Silver, 2006) (http://partsregistry.org/). The panoply of synthetic biological devices developed over the last decade has demonstrated that quantitative control over biological systems is possible in many contexts (Agapakis and Silver, 2009; Arkin and Fletcher, 2006; Boyle and Silver, 2009; Drubin et al., 2007; Endy, 2005; Haynes and Silver, 2009; Tyo et al., 2007).

Many synthetic biology endeavors also fall under the umbrella of metabolic engineering. Maximizing the production of a desired metabolite from a given feedstock mandates a quantitative evaluation and adjustment of cellular metabolism. To achieve this, synthetic biologists and metabolic engineers have sought fundamental engineering principles for biology. These principles have been inspired by traditional engineering disciplines as well as the unique properties of biological systems. In this review, we will explore both rational and evolutionary approaches to improving metabolic pathways.

2. Transcriptional and translational pathway control

A central challenge for every metabolic engineering project is to maximize product yields through pathway optimization (Keasling, 2010). Natural metabolic pathways are controlled by myriad regulatory systems, for example transcription factors and promoters that can be repurposed by synthetic biologists to modulate pathway components. Ideally, a quantitative understanding of the transcription, translation, interactions, and kinetics of a metabolic pathway as well as how that pathway interfaces with the host cell's metabolism enables the metabolic engineer to tune pathway components to maximize product yields. In practice, our ability to tune pathways has improved as the fundamental principles of metabolism and biological regulation continue to be discovered.

Many synthetic regulatory devices to date have leveraged elements of biology's "central dogma"—transcription, translation, as well as RNA processing—to modulate device behavior (Boyle and Silver, 2009). In the context of metabolic engineering, modifications to biological regulation are intended to maximize metabolic flux to the desired product. In most cases, this is accomplished via adjustments in enzyme expression levels, along

* Corresponding author at: Department of Systems Biology, Harvard Medical School, 200 Longwood Ave. WAB 536, Boston, MA 02115, USA. Fax: +1 617 432 5012.
E-mail address: pamela_silver@hms.harvard.edu (P.A. Silver).

1096-7176/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2011.10.003

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

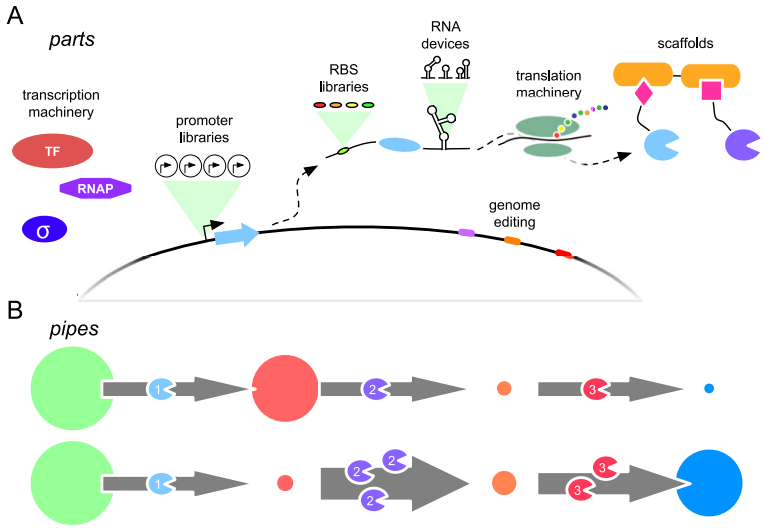


Fig. 1. Parts and pipes for the optimization of metabolic pathways. (A) Synthetic biologists use a variety of parts to adjust the functioning of metabolic pathways. Transcription machinery, enzyme promoters, ribosome binding sites (RBS), and translational machinery can be modified to adjust the concentration of an enzyme. RNA devices can modulate mRNA degradation and translation efficiency. Pathway enzymes can be assembled on scaffolds to optimize the spatial organization of a pathway. Genome editing approaches can be used to adjust host metabolism to improve flux through the target pathway. (B) A "pipe" of key pathway enzymes can be tuned to increase product titers. In this conceptual example, enzyme flux is represented by the size of the gray arrows. Metabolite concentrations are represented by the size of the circles between enzymes. In this example, increasing the concentration of the second and third enzymes in the pathway increases the titer of the product. Note that decreasing the concentration of intermediate metabolites can be beneficial; this is often the case when intermediates are harmful to the host cell. Increasing enzymes does not always improve product titers and can in fact be detrimental. In this review, we present synthetic biological parts that enable optimization of metabolic pipes.

with the elimination of competing pathways via gene knockout (Stephanopoulos, 1999).

The structure and function of evolved metabolic networks suggests that this process of pathway optimization requires an understanding of how control is distributed across the entire pathway (Dekel and Alon, 2005; Fell, 1997; Zaslaver et al., 2004). In essence, pathway optimization is a multivariate problem, with no single "rate limiting step" to target. Furthermore, simple overexpression of pathway enzymes is often detrimental to product yields, through both the depletion of essential cellular reserves and the buildup of toxic metabolic intermediates (Alper et al., 2005; Jones et al., 2000; Raab et al., 2005). Efforts to model synthetic biological circuits have also revealed that desired device behavior is highly dependent on the concentration of the device components within cells (Ajo-Franklin et al., 2007; Anderson et al., 2007; Elowitz and Leibler, 2000). As a consequence, methods for the control of protein expression levels are essential to metabolic engineering and synthetic biology in general.

2.1. The rational approach

Ideally, a quantitative understanding of the pathway to be engineered can allow metabolic engineers to determine optimal expression level of pathway elements *a priori*. Synthetic biologists have found that tight control of protein concentrations is required to achieve robust behavior of genetic circuits (Ajo-Franklin et al., 2007; Anderson et al., 2007; Boyle and Silver, 2009). Forward engineering of metabolic pathways can be facilitated by a variety of standardized and characterized control elements available to the metabolic engineering community.

For decades, promoter elements have been used to modify gene expression (Reznikoff et al., 1969). In recent years, a number of groups have assembled and characterized promoter libraries for common industrial hosts, such as *Escherichia coli*, *Saccharomyces cerevisiae*, and *Pichia pastoris* (Alper et al., 2005; Cox et al., 2007; Davis et al., 2010; Hartner et al., 2008; Nevoigt et al., 2006). In each case, native promoters were mutated or recombined to generate a group of promoters of varying strengths. Work has begun to develop standard metrics for promoter characterization, but remains dependent on high-throughput screening of promoter libraries rather than *in silico* prediction (Bayer, 2010; Kelly et al., 2009). This issue is compounded by the contextual variability of expression levels in response to environmental factors such as temperature or carbon source (Kelly et al., 2009).

Ribosome Binding Sites (RBS) mediate translation initiation, with variation in RBS sequence directly affecting translation efficiency. Thermodynamic models of translation initiation have been generated that now allow *a priori* design of RBS appropriate for a desired expression level. The RBS Calculator (http://salis.psu.edu/software/) generates a customized RBS for a given gene based on the desired translation initiation rate, gene sequence, and host organism. The RBS Calculator was successfully utilized to predict RBS combinations that would permit the desired operation of a synthetic AND gate (Salis et al., 2009), a device that is highly dependent on the expression levels of the inputs to produce AND gate output (Anderson et al., 2007).

Modification of RNA degradation rates can also control steady-state expression levels. In *S. cerevisiae*, the Rnt1p RNAse recognizes and cleaves a specific class of RNA hairpin (Lamontagne et al., 2003). When Rnt1p target hairpins are placed in the untranslated region (UTR) of an mRNA transcript, Rnt1p degradation lowers the effective expression level of the target gene. A library of variable Rnt1p target hairpins has been constructed that permits quantitative control of *S. cerevisiae* gene expression (Babiskin and Smolke, 2011). A subset of this library was inserted into the 3′ UTR of GFP, mCherry, and squalene synthase (*ERG9*).

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

The strong rank order correlation of expression level between the GFP and mCherry variants (ρ=0.848) and between the GFP and ERG9 variants (ρ=0.844) indicates that Rnt1p elements are largely modular (Babiskin and Smolke, 2011).

Constitutive promoters and RNA elements are useful for maintaining a steady-state expression level. In a non-steady-state environment, cells maintain homeostasis by reacting to environmental changes; endogenous metabolic pathways dynamically respond to changes in intracellular metabolite concentrations. Regulated gene expression (Beckwith, 1967; Jacob and Monod, 1961), RNA riboswitches (Mandal and Breaker, 2004), and allosteric control of enzyme activity (Monod et al., 1963) provide this control over a wide range of contexts and timescales. Designing similar dynamics into engineered pathways could improve the performance of engineered strains at industrial scales, where reactor conditions are not always uniform (Holtz and Keasling, 2010).

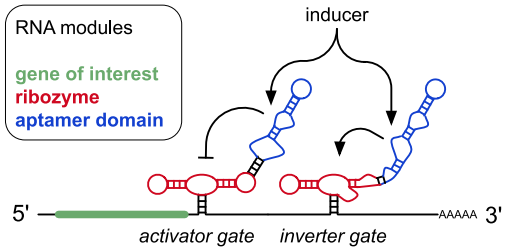
Modular RNA elements can be designed to provide a dynamic response to intracellular metabolite levels. Riboswitches are natural RNA elements that undergo a conformational change in response to a small-molecule ligand. When riboswitches are part of an mRNA molecule, this conformational change modulates the translation of the mRNA sequence. (Nahvi et al., 2002; Stoddard and Batey, 2006). Many riboswitches have been discovered in untranslated regions of mRNAs encoding for metabolic enzymes, offering a post-transcriptional layer of control over enzyme levels.

The potential for RNA-based multisite pathway modulation is exemplified in the 11 known S-adenosylmethionine (SAM) dependent riboswitches of *Bacillus subtilis*. In *B. subtilis*, much of the methionine biosynthesis pathway is regulated by SAM dependent riboswitches. These riboswitches function primarily through SAM-dependent conformational changes that trigger premature transcriptional termination, although a smaller subset disrupts translation initiation instead. Remarkably, each of the 11 riboswitches is independently tuned to a different SAM concentration. Furthermore, the termination efficiency of each SAM riboswitch in both the ligand bound and unbound conformations are different for each gene (Tomsic et al., 2008). Augmenting engineered metabolic pathways with small-molecule responsive RNA regulators could offer similarly distributed control (Beisel and Smolke, 2009).

A variety of synthetic RNA regulators have been designed to control gene expression. Synthetic RNA regulators can interact in *cis* with mRNA via aptamer domains to respond to small molecules (Bayer and Smolke, 2005; Win and Smolke, 2008), or make use of *trans*-acting RNA elements expressed off of an inducible promoter (Callura et al., 2010; Isaacs et al., 2004). Robust methods have been developed for the selection of RNA aptamer domains (Gilbert and Batey, 2005), and modular RNA elements can be combined to generate higher-order behaviors. For example, pairs of RNA aptamer domains alternately promoting or inhibiting translation of a transcript can serve as "bandpass filters," permitting mRNA translation between the range of concentrations set by the aptamer domains (Win and Smolke, 2008) (Fig. 2). Combining promoters and RBS tuned for steady-state performance with dynamically regulated RNA regulators may improve the robustness of engineered pathways.

2.2. The rationally irrational approach

Synthetic biologists strive to make biology "engineerable"—to discover modular biological elements that can be predictably assembled and designed to function robustly. Efforts to produce and characterize libraries of standardized parts have made progress towards this goal, yet the complexity of biological systems has kept biological engineering firmly in the trial and error stage.



inducer concentration	ribozyme activity	ribozyme activity	mRNA translation
low	on	off	off
medium	off	off	on
high	off	on	off

Fig. 2. An RNA-regulated bandpass filter. Two modular RNA regulators added to the 3′ untranslated region of an mRNA can be used to control mRNA translation in response to the concentration of a small molecule. Each regulator contains a self-cleaving ribozyme, coupled to an RNA aptamer domain. The activator gate ribozyme is repressed by the inducer, while the inverter gate ribozyme is activated by the same inducer. If the inducer concentration is between the ligand-binding thresholds of the two gates, both ribozymes are inactive and translation of the transcript is permitted (Win and Smolke, 2008).

Even synthetic devices with well-defined parameters for desired behavior require exhaustive characterization of the biological components to achieve functionality (Ajo-Franklin et al., 2007; Anderson et al., 2007). However, trial and error through the process of evolution has generated the biological diversity that synthetic biologists seek to redesign. In addition to traditional engineering principles, engineers of biological systems have access to the tools of selection and evolution; these tools can be leveraged to discover improvements to metabolic pathways. The ability to use these "rationally irrational" approaches is a core advantage to engineering biological systems.

Early metabolic engineering efforts relied on genomic mutagenesis to generate strains with desired properties (Stephanopoulos, 1999). If the phenotype of interest is accessible via a single mutation, mutagenesis is an acceptable approach. If the desired phenotype requires multiple mutations, however, the combinatorial expansion of the library size required to identify that phenotype makes untargeted mutagenesis practically infeasible (Dietrich et al., 2010b). Generating variation in a targeted subset of the genome enriches the resulting library for mutants with relevant phenotypes (Carr and Church, 2009).

Mutagenesis of the cellular transcriptional machinery can be used to adjust gene expression levels. In engineered cells, endogenous regulation often interferes with the functioning of heterologous pathways. Global Transcription Machinery Engineering (gTME) is an approach that modifies relative transcription rates across all genes simultaneously by selectively mutagenizing genes involved in the initiation of transcription. For example, mutagenesis of the *S. cerevisiae* TATA-binding protein *SPT15* and selection for improved ethanol tolerance yielded a mutant with a 20% higher biomass yield than the parent strain (Alper et al., 2006). GTME in *E. coli*, targeting the primary sigma factor σ^{70}, saw similar gains when applied to ethanol tolerance as well as 50% gains when applied to lycopene production (Alper and Stephanopoulos, 2007).

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

A more targeted approach to pathway adjustment is to selectively alter the regulation of pathway genes. Introducing RNAse cleavage sites or hairpin structures that alter mRNA stability into intergenic regions can result in different translation rates for two ORFs on the same mRNA (Smolke et al., 2000). Tunable Intergenic Regions (TIGR) are synthetic RNA constructs that include two hairpins joined by a RNAse cleavage site, and can be used to connect co-transcribed ORFs. Libraries of TIGR elements with a wide variety hairpin structures can be inserted between two co-transcribed genes to screen for optimal translation ratios (Pfleger et al., 2006).

TIGR elements have been used to improve the production of mevalonate in *E. coli* (Pfleger et al., 2006). This pathway requires the expression of acetoacetyl-CoA thiolase (AtoB), as well as the heterologous expression of hydroxy-methylglutaryl-CoA synthase (HMGS) and hydroxy-methylglutaryl-CoA reductase (HMGR). Inserting a library of TIGR elements into the AtoB-HMGS-HMGR operon identified a combination that increased mevalonate titers seven-fold over the initial AtoB-HMGS-HMGR operon. Each of the four best mevalonate producers identified in the screen lowered the expression levels of HMGS and HMGR relative to AtoB (Pfleger et al., 2006). HMG-CoA, the product of HMGS, was later shown to be cytotoxic to *E. coli* (Kizer et al., 2008); the best operons identified in the TIGR-mevalonate screen maintained or lowered HMG-CoA concentrations 11 h post-induction versus the parent strain (Pfleger et al., 2006).

New mutagenesis strategies are enabling iterative and simultaneous mutation of gene regulatory elements. Multiplex Automated Genome Engineering (MAGE) is a high throughput technique for the directed evolution of microbial genomes (Wang et al., 2009). MAGE combines both rational design and directed evolution approaches; specific genomic targets are selected for mutagenesis. For each genomic target, pools of degenerate oligonucleotides that retain homology to the target sequence are electroporated into the cells to be engineered. Multiple pools of oligonucleotides can be combined in a single electroporation step, allowing multiple genomic loci to be modified simultaneously. Iterative rounds of electroporation and growth generate a mixed population of cells with a variety of mutations at loci of interest (Fig. 3).

The utility of MAGE in improving pathway flux was evaluated in the context of engineered lycopene production in *E. coli*. Twenty endogenous genes known to affect lycopene yields were targeted with alternative RBS oligos. Simultaneously, four genes known to direct flux to competing pathways were targeted with oligos harboring nonsense mutations. Over 35 MAGE cycles, approximately 15 billion genetic variants were generated. Screening colonies based on the red pigmentation of lycopene identified a variant that produced fivefold more lycopene than the parent strain (Wang et al., 2009).

Combinatorial approaches are powerful tools for pathway optimization because they can adjust multiple gene levels simultaneously. Iterative pathway improvement, in which a single gene level is adjusted at a time, can fail to identify global maxima accessible by simultaneous perturbation (Alper and Stephanopoulos, 2007). Both gene knockout and upregulation studies have shown that mutations often interact in a cooperative and non-linear manner with regards to metabolite production (Kennedy et al., 2009). As a further complication, many modern metabolic engineering efforts involve the heterologous expression of enzymes from several different species in an unoptimized host (Agapakis et al., 2010; Bayer et al., 2009; Martin et al., 2003; Ro et al., 2006). Engineering these chimeric pathways to interface with host metabolism demands many factors be adjusted simultaneously.

Generating genomic or pathway-specific variation in gene regulation is only the first step in pathway optimization. Each approach outlined in this section was paired with a screening or selection strategy to identify improved product yields. Pathways that are not observable via high-throughput assays are less amenable to screening approaches. Selection strategies that connect pathway output to cell viability are designed ad hoc, and success is not guaranteed (Dietrich et al., 2010a). The lack of generalized methods for pathway screening and selection currently limits the broad application of combinatorial pathway optimization methods.

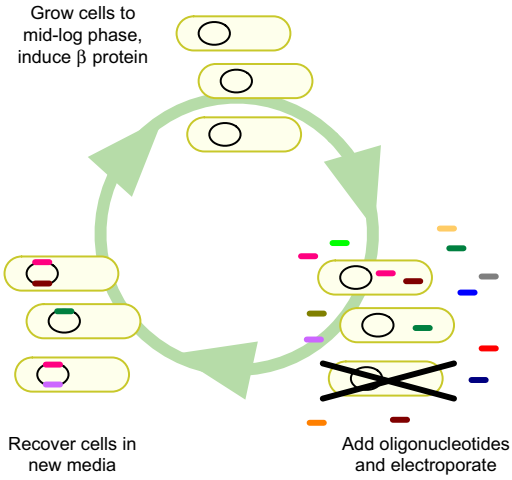


Fig. 3. The Multiplex Automated Genome Engineering (MAGE) cycle. MAGE incorporates oligonucleotides into *E. coli* by electroporation, with the λ-Red β protein integrating the oligonucleotides into the genome. Oligonucleotides can be synthesized to introduce mutations at precise genomic loci. Iterated rounds of MAGE introduce increasing amounts of diversity at these loci, although many cells are killed at the electroporation step (Wang et al., 2009).

3. Spatial pathway control

The spatial organization of cellular components is tightly controlled in all organisms, including prokaryotes. For example, many cyanobacteria target photosystems and electron transport machinery to a highly ordered thylakoid membrane (Nelson and Yocum, 2006), while maintaining carbon fixation enzymes in separate protein-bound compartments positioned along the cell axis (Savage et al., 2010) (Fig. 4C and D). Co-localization of pathway enzymes to the same subcellular organelle or compartment can increase the local concentration of pathway intermediates and exclude competing cytosolic pathways. Multienzyme complexes often arrange enzymes in defined stoichiometric ratios to improve enzyme saturation (Zhou et al., 2001). Direct linkage of enzyme active sites via substrate tunnels has also been observed in nature, such as in the synthesis of tryptophan (Hyde et al., 1988). The success of spatial pathway control in nature has led synthetic biologists to develop methods for adjusting the physical arrangement of metabolic pathways.

3.1. Scaffolds

Co-localization of related enzymes via direct linkage or scaffold proteins is an evolutionary development that has inspired new approaches to metabolic engineering. Polyketide synthases are modular enzymes that pass the growing polyketide chain from one enzymatic module to the next, much like an assembly line (Menzella et al., 2005). Cellulosomes are massive bacterial

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

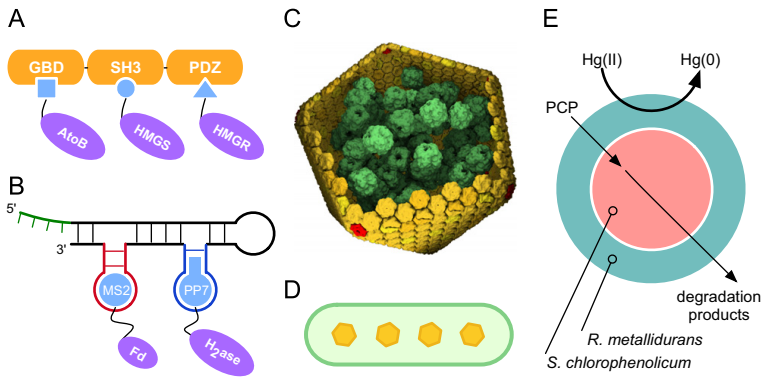


Fig. 4. Spatial optimization in natural and synthetic pathways across multiple scales. (A) At the protein scale, synthetic scaffolds can be used to bind mevalonate pathway enzymes in close proximity to each other. The scaffold contains binding domains for the GBD, SH3, and PDZ protein tags (Dueber et al., 2009). (B) Synthetic scaffolds also work for electron-transfer proteins, such as ferredoxin (Fd) and an [FeFe]-hydrogenase (H_2ase). In this example, an RNA scaffold was used to coordinate hydrogen production. RNA aptamers specific for the MS2 and PP7 protein tags allow control over enzyme binding. A 5′ extension (shown in green) allows attachment of this RNA building block onto extended scaffold architectures (Delebecque et al., 2011). (C) The carboxysome coordinates enzymes at a larger subcellular scale. RuBisCO (green) is tightly packed inside the icosahedral carboxysome shell (yellow). Carbonic anhydrase (not shown) provides RuBisCO with gaseous CO_2 inside the carboxysome (Savage et al., 2010). Carboxysome image used with permission from Bruno Afonso and David Savage. (D) At the cellular scale, carboxysomes in *S. elongatus* are evenly spaced across the length of the cell (Savage et al., 2010). (E) At the multicellular scale, microbial consortia can be assembled via microfluidic devices. Fibers of the PCP-degrading *S. chlorophenolicum* can be protected from Hg inhibition when coated with the Hg-reducing *R. metallidurans* (Kim et al., 2011).

complexes that arrange enzymes on scaffolds on the cell surface. The highly ordered arrangement of cellulosome enzymes facilitates the breakdown of plant cellulose, and work to heterologously express cellulosome assemblies has demonstrated that scaffolding increases enzyme activity in a cooperative manner (Gilbert, 2007; Mitsuzawa et al., 2009; Moraïs et al., 2010a, 2010b; Tsai et al., 2009).

Synthetic scaffolds have been constructed to improve the production of mevalonate (Dueber et al., 2009). As with the application of TIGR elements to the mevalonate pathway, the enzymes AtoB, HMGS, and HMGR were co-expressed in *E. coli*. Attachment of these enzymes to a scaffold protein was achieved by fusing the enzymes to the metazoan protein–protein interaction ligands GBD, SH3, and PDZ, respectively. Co-expressing a synthetic scaffold protein with cognate binding domains for the protein tags permitted all three enzymes to co-localize on the scaffold (Fig. 4A). Varying the number and order of binding sites for each tag allowed the enzyme stoichiometry on the scaffold to be tuned, as well as the relative positioning of each enzyme (Dueber et al., 2009).

Attaching the mevalonate pathway to the scaffold resulted in a striking 77-fold improvement in mevalonate yield over the unscaffolded pathway (Dueber et al., 2009). The scaffolding system appears to be generalizable to other metabolic pathways, as demonstrated by the use of the same scaffolding system for the production of glucaric acid in *E. coli*. Once again, the scaffold boosted product titers, with adjustments to scaffold binding site ratios increasing titers fivefold over the parent strain (Dueber et al., 2009; Moon et al., 2010).

The GBD/SH3/PDZ scaffolding system has been extended to a heterologous hydrogen production pathway (Agapakis et al., 2010). Pyruvate:ferredoxin oxidoreductase (PFOR) from *Desulfovibrio africanus* and ferredoxin and an [FeFe]-hydrogenase from *Clostridium acetobutylicum* were heterologously coexpressed with a scaffold. Unlike the mevalonate or glucaric acid pathways, the key intermediate for hydrogen production is not a small molecule. Instead, PFOR reduces the ferredoxin protein, which delivers the electrons to the hydrogenase. Despite this important difference, placing pathway components on a synthetic scaffold also resulted in increased product yields. The length of the linker between each component and the scaffold-binding ligand, the distance between binding sites on the scaffold, and the ratio and order of ligand binding sites all affected pathway yields. The study saw a threefold increase in hydrogen production with the best scaffold configuration over an unscaffolded parent strain. This work, along with evidence from natural and engineered signaling pathways (Bashor et al., 2008), demonstrate the utility of scaffolding in improving the specificity of protein–protein interactions.

Structural RNA and DNA devices are powerful alternatives to protein structures for spatially arranging biological parts. Tools for predicting secondary structures from the primary nucleic acid sequence are becoming sufficiently robust to enable the design of new RNA and DNA structures ab initio (Andronescu et al., 2004; Douglas et al., 2009). The field of DNA nanotechnology has demonstrated the versatility of DNA as a structural molecule (Aldaye et al., 2008; Shih and Lin, 2010), and RNA-based designs are beginning to be constructed in much the same manner (Chworos et al., 2004; Guo, 2010).

Functional in vivo RNA architectures were recently developed to scaffold a metabolic pathway. These RNA assemblies were used to coordinate the PFOR/ferredoxin/hydrogenase system in *E. coli*. Hydrogenase and ferredoxin proteins linked to PP7 and MS2 aptamer proteins were spatially organized onto an extended RNA scaffold bearing the corresponding RNA aptamers (Fig. 4B). As with the GBD/SH3/PDZ scaffold, hydrogen production benefited from scaffolding: the RNA scaffold increased hydrogen yield by 48-fold over the unscaffolded system (Delebecque et al., 2011). Since a vast array of multidimensional structures has already been designed with nucleic acids (Aldaye et al., 2008; Shih and Lin, 2010), there is great potential in the exploration of complex geometries for spatial optimization of metabolic pathways.

The exact mechanism behind yield improvements seen with scaffolds in various applications remains unclear (DeLisa and Conrado, 2009). Protein intermediates such as ferredoxin, which diffuse more slowly than small molecules, are likely to benefit from the local increase in concentration afforded by tethering to the scaffold (Agapakis et al., 2010; Cironi et al., 2008). For pathways with small molecule intermediates, it is more difficult

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

to prove that the local concentration of metabolic intermediates is increased near the scaffold. In the case of the mevalonate, the results are certainly indicative of increased pathway flux, as buildup of HMG-CoA is toxic to *E. coli* (Dueber et al., 2009). Importantly, synthetic scaffolds offer an elegant mechanism for balancing pathway flux through the precise adjustment of enzyme stoichiometries.

3.2. Subcellular compartments

Membrane- and protein-bound compartments encapsulate metabolic processes in prokaryotes and eukaryotes. Active transport or selective diffusion can boost substrate concentrations as well as protect a pathway from competing reactions. Compartments often host metabolic reactions that are harmful to the rest of the cell, or are thermodynamically infeasible in the cytoplasm (Chance et al., 1979; Feldman and Sigman, 1983; Page et al., 1998).

Many prokaryotes target critical metabolic reactions to protein-bound microcompartments. Perhaps the most well-studied of these compartments is the bacterial carboxysome, a structure that contains the key carbon fixation enzymes ribulose-1.5-bisphosphate carboxylase-oxygenase (RuBisCO) and carbonic anhydrase (Fig. 4C). The carboxysome is an evolutionary adaptation for a difficult metabolic task: to promote the fixation of gaseous carbon dioxide by RuBisCO while preventing oxygen from interfering with the reaction. The intricate organization of carboxysomes within *Synechococcus elongatus* suggests that the spatial positioning of carbon fixation is vital, even at the micron scale (Savage et al., 2010). Given the need for carbon-neutral or carbon-negative fuel sources (Savage et al., 2008), the bacterial carbon fixation machinery is a likely target for metabolic engineering.

Recent work has raised the possibility of repurposing prokaryotic microcompartments for metabolic engineering. The propanediol utilization (*pdu*) machinery of several bacterial species is enclosed in a proteinaceous shell much like a carboxysome (Parsons et al., 2008; Yeates et al., 2011). The *pdu* "metabolosome" of *Citrobacter freundii* was shown to confer the ability to metabolize propanediol when expressed heterologously in *E. coli*. Furthermore, electron micrographs appear to show successful assembly of the *pdu* shell proteins to form microcompartments in *E. coli* (Parsons et al., 2008). Intriguingly, there is evidence that the N-terminal domains of non-shell *pdu* proteins can be appropriated to target heterologous proteins to the metabolosome interior (Fan et al., 2010; Parsons et al., 2010). The pairing of microcompartment shells with novel biosynthetic pathways may expand the reach of bacterial metabolic engineering.

In eukaryotes, methods have been developed for the targeting of heterologous proteins to many membrane-bound organelles (Hood and Silver, 1999; Léon et al., 2006; Soll and Schleiff, 2004; Truscott et al., 2003). Efforts to synthesize methyl halides in *S. cerevisiae* have demonstrated the utility of localizing exogenous enzymes in appropriate subcellular environments. Researchers noted that the primary substrates of methyl halide transferases (MHT), SAM, and halide ions are sequestered in the yeast vacuole. Targeting the MHT from *Batis maritima* to the vacuole increased yields of methyl iodide by nearly 50 mg/L-h over targeting the identical MHT to the cytosol (Bayer et al., 2009). The benefits of compartmentalization are likely to be even greater for the heterologous expression of biosynthetic pathways that require an organelle, such as penicillin synthesis (Gidijala et al., 2009; Meijer et al., 2010).

3.3. Microbial consortia

The natural world has demonstrated that consortia and communities of organisms are capable of performing metabolic conversions that are difficult or thermodynamically unfavorable to do in a single cell (Wintermute and Silver, 2010a). From an engineering perspective, co-culture offers many of the same advantages as subcellular compartmentalization: incompatible metabolic reactions can be conducted in separate organelles or cells. The bovine rumen, which is itself compartmentalized, harbors a rich assortment of microorganisms that together metabolize cellulose (Annison and Bryden, 1998; Hungate, 1947; McAllister et al., 1994). Even communities of soil bacteria appear to be highly ordered (Young and Crawford, 2004). The success of microbial cooperation in the natural world has inspired efforts to engineer synthetic microbial consortia (Brenner et al., 2007, 2008; Eiteman et al., 2008; Kim et al., 2008; Shou et al., 2007; Wintermute and Silver, 2010b).

In recent years, efforts to model and construct stable engineered co-cultures have intensified in the systems and synthetic biology communities. Synthetic consortia have been established through the exchange of signaling molecules governing quorum-sensing circuits (Brenner et al., 2007) as well as the exchange of essential metabolites among complementary auxotrophs (Shou et al., 2007; Wintermute and Silver, 2010b). Constraint-based modeling frameworks familiar to metabolic engineers have been adapted to multi-organism systems, facilitating the development of compatible synthetic consortia (Klitgord and Segrè, 2010; Wintermute and Silver, 2010b).

As with proteins, the spatial arrangement of microbial consortia can be optimized for greater product yields. In a unique approach, wild-type *Sphingobium chlorophenolicum* and *Ralstonia metallidurans* were used as modular parts to assemble a structure capable of degrading pentachlorophenol (PCP) in the presence of Hg(II) (Kim et al., 2011) (Fig. 4E). PCP and Hg(II) are particularly harsh industrial pollutants that are often produced together. While *S. chlorophenolicum* is capable of degrading PCP, it is inhibited by Hg(II). *R. metallidurans*, a mercuric ion reducer, can reduce Hg(II) to Hg(0). Microfluidic laminar flow devices were used to assemble fibers of *S. chlorophenolicum* wrapped in a protective shell of *R. metallidurans*. The hybrid fibers were capable of fully degrading 120 μM PCP in the presence of micromolar Hg(II). A well-mixed solution of both species was essentially incapable of PCP degradation.

Synthetic microbial consortia consisting of wild-type organisms may be well suited for bioremediation and other ecologically sensitive applications. Wild-type organisms are not subject to laws and regulations concerning the deployment of genetically modified organisms into the environment. Ideally, spatially optimized consortia assembled of species native to polluted areas could be administered to accelerate the bioremediation process. The possibility of engineered microbes simultaneously obsolescing harsh industrial processes while cleaning up existing industrial pollution is an attractive vision for the future.

4. Modeling and measuring the metabolic network

In many ways, metabolites are the "dark matter" of the cell—their existence, intracellular concentrations, and fluxes are difficult to derive from genomic information and difficult to experimentally measure (Blow, 2008). Combining tools for characterizing the status of the metabolome with robust metabolic models is a foundational mission for systems and synthetic biology.

Genome-scale constraint-based models of cellular metabolism have been invaluable tools for *in silico* screening of mutant backgrounds suitable for metabolic engineering. These models incorporate the stoichiometry of all known metabolic reactions in a cell subject to linear or quadratic constraints, such as mass

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

balance, media composition, and thermodynamic limits on flux direction (Schellenberger et al., 2011; Segrè et al., 2002; Varma and Palsson, 1994). The result is a steady-state approximation of metabolic fluxes. Metabolic reactions can be added and removed to model heterologous gene expression and gene knockouts, respectively. Models of this type have successfully been employed to identify genetic modifications to improve product yields (Bro et al., 2006; Brochado et al., 2010; Burgard et al., 2003; Kennedy et al., 2009; Pharkya et al., 2004).

In order to model the entire metabolic network, constraint-based models eschew kinetic parameters and do not predict metabolite concentrations (Varma and Palsson, 1994). More quantitative approaches such as Metabolic Control Analysis are also more dependent on experimentally determined parameters; errors in parameter estimation become increasingly problematic as the model size increases (Schuster, 1999). Thus, constraint-based models are preferable for the forward engineering of the metabolic network, while detailed kinetic models of select pathways are beneficial for pathway tuning. As with all biological modeling, experimental analysis is crucial for evaluating the predictive value of metabolic models in a given context.

Multiple studies have attempted to reconcile transcriptome data with proteomic or metabolomic measurements in engineered cells (Bradley et al., 2009; Fendt et al., 2010; Ishii et al., 2007; Moxley et al., 2009). The integration of -omics level data with network-scale metabolic models benefits both *a priori* prediction as well as *post hoc* evaluation of metabolic engineering. In particular, developing quantitative models for the relationship between transcript levels and metabolite pools and fluxes would allow metabolomic data to be inferred from the vast number of microarray datasets that are already available (Yizhak et al., 2010). The chemically uniform nature of mRNA transcripts allows the reliable collection of total mRNA in a single extraction condition. Due to the chemical diversity of small molecules within cells, extraction conditions limit the extent of the metabolome that is observed (Yanes et al., 2011).

Coordination between gene expression and metabolite concentrations appears to be dependent on the type of perturbation. Comparison of the transcriptome and metabolome in *E. coli* over a range of growth rates revealed that enzyme transcript and protein levels increased with increasing growth rates while metabolite pools remained steady (Ishii et al., 2007). It was also noted that gene deletions that reverse the flux direction of the pentose phosphate pathway did not significantly alter enzyme levels or metabolite pool sizes. In addition, it was observed that metabolic enzymes do not appreciably up-regulate to compensate for enzyme knockouts (Ishii et al., 2007). Another study comparing the transcriptome and metabolome of *S. cerevisiae* during carbon and nitrogen starvation observed coordinated changes in expression and metabolite levels. In this study, a network model identified novel coordinated gene-metabolite pairs in this dataset (Bradley et al., 2009).

Further work is required to identify contexts in which transcript levels correlate to metabolite concentrations. Comparison of the above studies suggests a differential metabolic response to enzyme knockouts versus shifting media conditions (Bradley et al., 2009; Ishii et al., 2007); this could be a consequence of evolutionary selection for robustness against condition changes (Cornelius et al., 2011; Segrè et al., 2002). Alternatively, *E. coli* and *S. cerevisiae* may simply respond differently to metabolic perturbations. A significant confounding issue is that major metabolic flux alterations can occur without major shifts in enzyme or metabolite concentrations (Fell, 1997; Ishii et al., 2007).

Two recent studies measured transcriptomic and metabolomic shifts in *S. cerevisiae* in response to the deletion of global regulatory genes rather than enzymes (Fendt et al., 2010; Moxley et al., 2009). Following the deletion of the Gcn4p, a global stress response regulator, it was observed that flux control was highest among metabolites that were involved in many enzymatic reactions (Moxley et al., 2009). This raises the possibility of utilizing network topology to inform metabolic engineering. In the case of central carbon metabolism, deletion of the glycolysis-activating transcription factor Gcr2p showed a negative correlation between enzyme levels and associated metabolite levels (Fendt et al., 2010). This could be indicative of a buffering phenomenon, in which changes in metabolite pools counteract enzyme concentration changes to maintain a steady pathway flux.

A grand unifying theory of metabolism has not yet arisen from these meta-omics studies. It is possible that a truly general relationship between gene expression and metabolic concentrations does not exist but several important observations about their correlation have been made. Overall, it appears that evolved metabolic networks are quite robust in response to genetic and environmental perturbations. This is corroborated by many of the metabolic engineering efforts that we have reviewed, in which multiple perturbations were required to improve product yields. This could also explain why conservative modeling frameworks such as Minimization of Metabolic Adjustment (MOMA) are often more accurate than models that assume that mutant flux is optimized for maximum growth. Instead, MOMA-derived solutions assume that mutant fluxes are regulated by the cell to approximate the wild-type flux distribution (Segrè et al., 2002). Integrative data from –omics scale datasets may help to identify genes that disrupt the resistance to perturbations in pathways of interest.

5. Conclusions

As a practical application of synthetic biology, metabolic engineering has field-tested emerging biological design principles. Through these efforts, it has become increasingly apparent that rational design approaches are limited by our understanding of biological systems. The complexity of living cells far surpasses the complexity of human-made devices. The tremendous improvement in DNA sequencing and assembly techniques is now bringing about an era in which cells themselves can be man-made devices (Benders et al., 2010; Gibson et al., 2010; Lartigue et al., 2009), but our ability to modify cells has outpaced our ability to predict how those modifications will function. Fortunately, the ability of living systems to self-replicate has led animal and plant breeders, geneticists, molecular biologists, metabolic engineers, and now synthetic biologists to utilize selective pressure in their research.

Incorporating irrational design into synthetic biology does not require an abandonment of forward engineering approaches. Instead, the emerging engineering design cycle of synthetic biology and metabolic engineering appears to include both *in silico* modeling and prediction as well as directed evolution and screening. Similarly, experimental analysis remains invaluable for hypothesis generation as well as confirmation. In our own work, we utilized FBA to predict knockout combinations in *S. cerevisiae* likely to increase production of formic acid. Our initial strain produced formic acid as predicted, but at modest levels. Expression analysis and metabolic phenotyping allowed us to identify further genetic interventions that boosted formic acid titers (Kennedy et al., 2009). In the case of the mevalonate-based isoprenoid pathway, transcriptomic and metabolomic methods revealed that the cytotoxicity of HMG-CoA was causing stress responses that negatively impacted isoprenoid titers (Kizer et al., 2008; Martin et al., 2003). Based on these experiments, researchers were able to counteract the effect by varying pathway

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

expression levels (Pfleger et al., 2006), by the addition of palmitic acid (Kizer et al., 2008), and by the use of engineered scaffolds (Dueber et al., 2009). Undoubtedly genome-scale optimization approaches such as gTME and MAGE can identify further regulatory adjustments that have not been predicted.

The ability to synthesize any DNA sequence has increased the reach of metabolic engineering. Modern metabolic engineering efforts often assemble parts from disparate species into novel pathways. Recent work to produce fatty acids in *E. coli* utilized heterologous genes from *S. cerevisiae*, *M. musculus*, *A. calcoaceticus*, *A. baylyi ADP1*, *Z. mobilis ZM4*, *C. stercorarium*, *B. ovatus*, *U. california*, *C. hookeriania*, and *A. thaliana*, sampling multiple expression level combinations (Steen et al., 2010). Low-cost synthesis will undoubtedly drive more "synthetic metagenomics" studies, in which libraries of homologous enzymes are synthesized and evaluated (Bayer et al., 2009). Demands for renewable energy have led synthetic biologists to construct more spatially complex pathways, such as the expression of the electron transfer apparatus in *E. coli* (Jensen et al., 2010). Exploring all relevant expression level variants and spatial arrangements becomes increasingly difficult in these complex pathways.

The engineering of complex pathways can be divided into two stages: a proof of concept stage where novel enzyme combinations are determined to produce a desired product, and an optimization stage where regulatory adjustments are made to improve product yields. As engineered biosynthetic pathways become more complex, they also become more difficult to tune. New biofabrication facilities such as the BIOFAB (http://www.biofab.org/) seek to abstract pathway optimization from pathway design by developing rapid prototyping services. The BIOFAB is intended to provide new libraries of characterized regulatory elements, as well as facilitate prototyping of collaborator's synthetic devices via high-throughput cloning and testing (Bayer, 2010). Applying forward engineering and directed evolution approaches in a high-throughput manner will generate parts and data that will improve our ability to rationally engineer cells.

The growth of synthetic biology has often been compared to the personal computer revolution of the late twentieth century. The home personal computer was made possible by the availability of high-quality off-the-shelf electronic components that could be assembled by technically inclined enthusiasts. Molecular biologists have generated a vast assortment of biological parts, but inadequate characterization has limited their general usefulness. In many ways, the current stage of synthetic biology is more analogous to the early days of heavier-than-air flight. Despite our current sophistication regarding the computational design of aircraft, flight principles were first elucidated through exhaustive wind tunnel experiments and test flights (Carlson, 2010). Similarly, the successes and failures of synthetic biology continue to reveal biological design principles.

Funding

P.M.B is supported by the Harvard University Center for the Environment and the NSF Synthetic Biology Engineering Research Center. P.A.S is supported by the Department of Defense (DOD W81XWH-08-1-0608) and the National Institutes of Health (NIHGM36373-22).

Acknowledgments

We would like to thank Christina Agapakis, Camille Delebecque, and Jake Wintermute for their critical reading of the manuscript. We would also like to thank Christian Boehm for assistance with the manuscript.

References

Agapakis, C.M., Ducat, D.C., Boyle, P.M., Wintermute, E.H., Way, J.C., Silver, P.A., 2010. Insulation of a synthetic hydrogen metabolism circuit in bacteria. J. Biol. Eng. 4, 3.

Agapakis, C.M., Silver, P.A., 2009. Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks. Mol. Biosyst. 5, 704–713.

Ajo-Franklin, C.M., Drubin, D.A., Eskin, J.A., Gee, E.P., Landgraf, D., Phillips, I., Silver, P.A., 2007. Rational design of memory in eukaryotic cells. Genes Dev. 21, 2271–2276.

Aldaye, F.A., Palmer, A.L., Sleiman, H.F., 2008. Assembling materials with DNA as the guide. Science (New York, NY) 321, 1795–1799.

Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., 2005. Tuning genetic control through promoter engineering. Proc. Natl. Acad. Sci. USA 102, 12678–12683.

Alper, H., Moxley, J., Nevoigt, E., Fink, G.R., Stephanopoulos, G., 2006. Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314, 1565–1568.

Alper, H., Stephanopoulos, G., 2007. Global transcription machinery engineering: a new approach for improving cellular phenotype. Metab. Eng. 9, 258–267.

Anderson, J.C., Dueber, J.E., Leguia, M., Wu, G.C., Goler, J.A., Arkin, A.P., Keasling, J.D., 2010. BglBricks: A flexible standard for biological part assembly. J. Biol. Eng. 4, 1.

Anderson, J.C., Voigt, C., Arkin, A.P., 2007. Environmental signal integration by a modular AND gate. Mol. Syst. Biol. 3, 133.

Andronescu, M., Fejes, A.P., Hutter, F., Hoos, H.H., Condon, A., 2004. A new algorithm for RNA secondary structure design. J. Mol. Biol. 336, 607–624.

Annison, E.F., Bryden, W.L., 1998. Perspectives on ruminant nutrition and metabolism I. Metabolism in the rumen. Nutr. Res. Rev. 11, 173–198.

Arkin, A.P., Fletcher, D.A., 2006. Fast, cheap and somewhat in control. Genome Biol. 7, 114.

Babiskin, A.H., Smolke, C.D., 2011. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. Mol. Syst. Biol. 7, 471.

Bashor, C.J., Helman, N.C., Yan, S., Lim, W.A., 2008. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. Science 319, 1539–1543.

Bayer, T.S., 2010. Grand challenge commentary: transforming biosynthesis into an information science. Nat. Chem. Biol. 6, 859–861.

Bayer, T.S., Smolke, C.D., 2005. Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat. Biotechnol. 23, 337–343.

Bayer, T.S., Widmaier, D.M., Temme, K., Mirsky, E.A., Santi, D.V., Voigt, C.A., 2009. Synthesis of methyl halides from biomass using engineered microbes. J. Am. Chem. Soc. 131, 6508–6515.

Beckwith, J.R., 1967. Regulation of the lac operon. Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. Science 156, 597–604.

Beisel, C.L., Smolke, C.D., 2009. Design principles for riboswitch function. PLoS Comput. Biol. 5, e1000363.

Benders, G.A., Noskov, V.N., Denisova, E.A., Lartigue, C., Gibson, D.G., Assad-Garcia, N., Chuang, R.-Y., Carrera, W., Moodie, M., Algire, M.A., Phan, Q., Alperovich, N., Vashee, S., Merryman, C., Venter, J.C., Smith, H.O., Glass, J.I., Hutchison, C.A., Cloning whole bacterial genomes in yeast. Nucl. Acids Res., 2010.

Blow, N., 2008. Metabolomics. biochemistry's new look. Nature 455, 697–700.

Boyle, P.M., Silver, P.A., 2009. Harnessing nature's toolbox: regulatory elements for synthetic biology. J. Roy. Soc., Interface/Roy. Soc. 6 (Suppl. 4), S535–S546.

Bradley, P.H., Brauer, M.J., Rabinowitz, J.D., Troyanskaya, O.G., 2009. Coordinated concentration changes of transcripts and metabolites in *Saccharomyces cerevisiae*. PLoS Comput. Biol. 5, e1000270.

Brenner, K., Karig, D.K., Weiss, R., Arnold, F.H., 2007. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. Proc. Natl. Acad. Sci. USA 104, 17300–17304.

Brenner, K., You, L., Arnold, F.H., 2008. Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol. 26, 483–489.

Bro, C., Regenberg, B., Förster, J., Nielsen, J., 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. Metab. Eng. 8, 102–111.

Brochado, A.R., Matos, C., Møller, B.L., Hansen, J., Mortensen, U.H., Patil, K.R., 2010. Improved vanillin production in baker's yeast through in silico design. Microb. Cell Factor. 9, 84.

Burgard, A., Pharkya, P., Maranas, C., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–657.

Callura, J.M., Dwyer, D.J., Isaacs, F.J., Cantor, C.R., Collins, J.J., 2010. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. Proc. Natl. Acad. Sci. USA 107, 15898–15903.

Carlson, R.H., 2010. Biology is Technology: The Promise, Peril, and New Business of Engineering Life. Harvard University Press, Cambridge, MA.

Carr, P.A., Church, G.M., 2009. Genome engineering. Nat. Biotechnol. 27, 1151–1162.

Chance, B., Sies, H., Boveris, A., 1979. Hydroperoxide metabolism in mammalian organs. Physiol. Rev. 59, 527–605.

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

Chworos, A., Severcan, I., Koyfman, A.Y., Weinkam, P., Oroudjev, E., Hansma, H.G., Jaeger, L., 2004. Building programmable jigsaw puzzles with RNA. Science 306, 2068–2072.
Cironi, P., Swinburne, I.A., Silver, P.A., 2008. Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. J. Biol. Chem. 283, 8469–8476.
Cornelius, S.P., Lee, J.S., Motter, A.E., 2011. Dispensability of *Escherichia coli*'s latent pathways. Proc. Natl. Acad. Sci. USA 108, 3124–3129.
Cox, R.S., Surette, M.G., Elowitz, M.B., 2007. Programming gene expression with combinatorial promoters. Mol. Syst. Biol. 3, 145.
Davis, J.H., Rubin, A.J., Sauer, R.T., Design, construction and characterization of a set of insulated bacterial promoters. Nucl. Acids Res., 2010.
Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436, 588–592.
Delebecque, C.J., Lindner, A.B., Silver, P.A., Aldaye, F.A., 2011. Organization of intracellular reactions with rationally designed RNA assemblies. Science (New York, NY) 333, 470–474.
DeLisa, M.P., Conrado, R.J., 2009. Synthetic metabolic pipelines. Nat. Biotechnol. 27, 728–729.
Dietrich, J.A., McKee, A.E., Keasling, J.D., 2010a. High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu. Rev. Biochem. 79, 563–590.
Dietrich, J.A., Mckee, A.E., Keasling, J.D., 2010b. High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu. Rev. Biochem. 79, 563–590.
Douglas, S.M., Marblestone, A.H., Teerapittayanon, S., Vazquez, A., Church, G.M., Shih, W.M., 2009. Rapid prototyping of 3D DNA-origami shapes with caDNAno. Nucl. Acids Res. 37, 5001–5006.
Drubin, D.A., Way, J.C., Silver, P.A., 2007. Designing biological systems. Genes Dev. 21, 242–254.
Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L.J., Keasling, J.D., 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. 27, 753–759.
Eiteman, M.A., Lee, S.A., Altman, E., 2008. A co-fermentation strategy to consume sugar mixtures effectively. J. Biol. Eng. 2, 3.
Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338.
Endy, D., 2005. Foundations for engineering biology. Nature 438, 449–453.
Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., Yeates, T.O., Bobik, T.A., 2010. Short N-terminal sequences package proteins into bacterial microcompartments. Proc. Natl. Acad. Sci. USA 107, 7509–7514.
Feldman, R.I., Sigman, D.S., 1983. The synthesis of ATP by the membrane-bound ATP synthase complex from medium 32Pi under completely uncoupled conditions. J. Biol. Chem. 258, 12178–12183.
Fell, D.A., 1997. Understanding Control of Metabolism. Portland Press, London, UK.
Fendt, S.-M., Buescher, J.M., Rudroff, F., Picotti, P., Zamboni, N., Sauer, U., 2010. Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. Mol. Syst. Biol. 6, 356.
Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi, Z.-Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A., Smith, H.O., Venter, J.C., Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010.
Gidijala, L., Kiel, J.A.K.W., Douma, R.D., Seifar, R.M., van Gulik, W.M., Bovenberg, R.A.L., Veenhuis, M., van der Klei, I.J., 2009. An engineered yeast efficiently secreting penicillin. PLoS ONE 4, e8317.
Gilbert, H.J., 2007. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. Mol. Microbiol. 63, 1568–1576.
Gilbert, S.D., Batey, R.T., 2005. Riboswitches: natural SELEXion. Cell Mol. Life Sci. 62, 2401–2404.
Guo, P., 2010. The emerging field of RNA nanotechnology. Nat. Nanotechnol. 5, 833–842.
Hartner, F.S., Ruth, C., Langenegger, D., Johnson, S.N., Hyka, P., Lin-Cereghino, G.P., Lin-Cereghino, J., Kovar, K., Cregg, J.M., Glieder, A., 2008. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucl. Acids Res. 36, e76.
Haynes, K.A., Silver, P.A., 2009. Eukaryotic systems broaden the scope of synthetic biology. J. Cell Biol. 187, 589–596.
Holtz, W.J., Keasling, J.D., 2010. Engineering static and dynamic control of synthetic pathways. Cell 140, 19–23.
Hood, J.K., Silver, P.A., 1999. In or out? Regulating nuclear transport. Curr. Opin. Cell Biol. 11, 241–247.
Hungate, R.E., 1947. Studies on cellulose fermentation: III. The culture and isolation for cellulose-decomposing bacteria from the rumen of cattle. J. Bacteriol. 53, 631–645.
Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., Davies, D.R., 1988. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. J. Biol. Chem. 263, 17857–17871.
Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R., Collins, J.J., 2004. Engineered riboregulators enable post-transcriptional control of gene expression. Nat. Biotechnol. 22, 841–847.
Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P.Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., Tomita, M., 2007. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 316, 593–597.
Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.
Jensen, H.M., Albers, A.E., Malley, K.R., Londer, Y.Y., Cohen, B.E., Helms, B.A., Weigele, P., Groves, J.T., Ajo-Franklin, C.M., 2010. Engineering of a synthetic electron conduit in living cells. Proc. Natl. Acad. Sci. USA.
Jones, K.L., Kim, S.W., Keasling, J.D., 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. Metab. Eng. 2, 328–338.
Keasling, J.D., 2010. Manufacturing molecules through metabolic engineering. Science 330, 1355–1358.
Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., De Mora, K., Glieberman, A.L., Monie, D.D., Endy, D., 3, 2009. Measuring the activity of BioBrick promoters using an in vivo reference standard. J. Biol. Eng., 4.
Kennedy, C.J., Boyle, P.M., Waks, Z., Silver, P.A., 2009. Systems-level engineering of nonfermentative metabolism in yeast. Genetics 183, 385–397.
Kim, H.J., Boedicker, J.Q., Choi, J.W., Ismagilov, R.F., 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. Proc. Natl. Acad. Sci. USA 105, 18188–18193.
Kim, H.J., Du, W., Ismagilov, R.F., 2011. Complex function by design using spatially pre-structured synthetic microbial communities: degradation of pentachlorophenol in the presence of Hg(ii). Integr. Biol. (Camb.) 3, 126–133.
Kizer, L., Pitera, D.J., Pfleger, B.F., Keasling, J.D., 2008. Application of functional genomics to pathway optimization for increased isoprenoid production. Appl. Environ. Microbiol. 74, 3229–3241.
Klitgord, N., Segrè, D., 2010. Environments that induce synthetic microbial ecosystems. PLoS Comput. Biol. 6, e1001002.
Knight, T., 2003. Idempotent vector design for standard assembly of biobricks. DSpace. MIT Artificial Intelligence Laboratory; MIT Synthetic Biology Working Group.
Lamontagne, B., Ghazal, G., Lebars, I., Yoshizawa, S., Fourmy, D., Elela, S.A., 2003. Sequence dependence of substrate recognition and cleavage by yeast RNase III. J. Mol. Biol. 327, 985–1000.
Lartigue, C., Vashee, S., Algire, M., Chuang, R., Benders, G., Ma, L., Noskov, V., Denisova, E., Gibson, D., Assad-Garcia, N., Alperovich, N., Thomas, D., Merryman, C., Hutchison, C., Smith, H., Venter, J., Glass, J., Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science, 2009.
Léon, S., Goodman, J.M., Subramani, S., 2006. Uniqueness of the mechanism of protein import into the peroxisome matrix: transport of folded, co-factor-bound and oligomeric proteins by shuttling receptors. Biochim. Biophys. Acta 1763, 1552–1564.
Mandal, M., Breaker, R.R., 2004. Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol. 5, 451–463.
Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D., Keasling, J.D., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechnol. 21, 796–802.
McAllister, T.A., Bae, H.D., Jones, G.A., Cheng, K.J., 1994. Microbial attachment and feed digestion in the rumen. J. Anim. Sci. 72, 3004–3018.
Meijer, W.H., Gidijala, L., Fekken, S., Kiel, J.A.K.W., van den Berg, M.A., Lascaris, R., Bovenberg, R.A.L., van der Klei, I.J., 2010. Peroxisomes are required for efficient penicillin biosynthesis in *Penicillium chrysogenum*. Appl. Environ. Microbiol. 76, 5702–5709.
Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A., Santi, D.V., 2005. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat. Biotechnol. 23, 1171–1176.
Mitsuzawa, S., Kagawa, H., Li, Y., Chan, S.L., Paavola, C.D., Trent, J.D., 2009. The rosettazyme: a synthetic cellulosome. J. Biotechnol. 143, 139–144.
Monod, J., Changeux, J.P., Jacob, F., 1963. Allosteric proteins and cellular control systems. J. Mol. Biol. 6, 306–329.
Moon, T.S., Dueber, J.E., Shiue, E., Prather, K.L.J., 2010. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. Metab. Eng. 12, 298–305.
Moraïs, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D.B., Bayer, E.A., 2010a. Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. mBio, 1.
Moraïs, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., Wilson, D.B., Bayer, E.A., 2010b. Contribution of a xylan-binding module to the degradation of a complex cellulosic substrate by designer cellulosomes. Appl. Environ. Microbiol. 76, 3787–3796.
Moxley, J.F., Jewett, M.C., Antoniewicz, M.R., Villas-Boas, S.G., Alper, H., Wheeler, R.T., Tong, L., Hinnebusch, A.G., Ideker, T., Nielsen, J., Stephanopoulos, G., 2009. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. Proc. Natl Acad. Sci. USA 106, 6477–6482.
Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., Breaker, R.R., 2002. Genetic control by a metabolite binding mRNA. Chem. Biol. 9, 1043.
Nelson, N., Yocum, C.F., 2006. Structure and function of photosystems I and II. Annu. Rev. Plant Biol. 57, 521–565.
Nevoigt, E., Kohnke, J., Fischer, C.R., Alper, H., Stahl, U., Stephanopoulos, G., 2006. Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 72, 5266–5273.
Page, L.J., Darmon, A.J., Uellner, R., Griffiths, G.M., 1998. L is for lytic granules: lysosomes that kill. Biochim. Biophys. Acta 1401, 146–156.

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

Parsons, J.B., Dinesh, S.D., Deery, E., Leech, H.K., Brindley, A.A., Heldt, D., Frank, S., Smales, C.M., Lünsdorf, H., Rambach, A., Gass, M.H., Bleloch, A., McClean, K.J., Munro, A.W., Rigby, S.E.J., Warren, M.J., Prentice, M.B., 2008. Biochemical and structural insights into bacterial organelle form and biogenesis. J. Biol. Chem. 283, 14366–14375.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., Warren, M.J., 2010. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol. Cell 38, 305–315.

Pfleger, B.F., Pitera, D.J., Smolke, C.D., Keasling, J.D., 2006. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat. Biotechnol. 24, 1027–1032.

Pharkya, P., Burgard, A.P., Maranas, C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367–2376.

Phillips, I., Silver, P., 2006. A new biobrick assembly strategy designed for facile protein engineering. DSpace. MIT Artificial Intelligence Laboratory: MIT Synthetic Biology Working Group, Massachusetts Institute of Technology, Cambridge, MA 〈http://hdl.handle.net/1721.1/32535〉.

Raab, R.M., Tyo, K., Stephanopoulos, G., 2005. Metabolic engineering. Adv. Biochem. Eng. Biotechnol. 100, 1–17.

Reznikoff, W.S., Miller, J.H., Scaife, J.G., Beckwith, J.R., 1969. A mechanism for repressor action. J. Mol. Biol. 43, 201–213.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R., Keasling, J.D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature, 440, 2006, 940.

Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. 27, 946–950.

Savage, D., Afonso, B., Chen, A., Silver, P., 2010. Spatially ordered dynamics of the bacterial carbon fixation machinery. Science 327, 1258.

Savage, D.F., Way, J., Silver, P.A., 2008. Defossiling fuel: how synthetic biology can transform biofuel production. ACS Chem. Biol. 3, 13–16.

Schellenberger, J., Lewis, N.E., Palsson, B.Ø., 2011. Elimination of thermodynamically infeasible loops in steady-state metabolic models. Biophys. J. 100, 544–553.

Schuster, S., 1999. Use and limitations of modular metabolic control analysis in medicine and biotechnology. Metab. Eng. 1, 232–242.

Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99, 15112–15117.

Shih, W.M., Lin, C., 2010. Knitting complex weaves with DNA origami. Curr. Opin. Struct. Biol. 20, 276–282.

Shou, W., Ram, S., Vilar, J.M.G., 2007. Synthetic cooperation in engineered yeast populations. Proc. Natl. Acad. Sci. USA 104, 1877–1882.

Smolke, C.D., Carrier, T.A., Keasling, J.D., 2000. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. Appl. Environ. Microbiol. 66, 5399–5405.

Soll, J., Schleiff, E., 2004. Protein import into chloroplasts. Nat. Rev. Mol. Cell Biol. 5, 198–208.

Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B., Keasling, J.D., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463, 559–562.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. Metab. Eng. 1, 1–11.

Stoddard, C.D., Batey, R.T., 2006. Mix-and-match riboswitches. ACS Chem. Biol. 1, 751–754.

Tomsic, J., McDaniel, B.A., Grundy, F.J., Henkin, T.M., 2008. Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. J. Bacteriol. 190, 823–833.

Truscott, K.N., Brandner, K., Pfanner, N., 2003. Mechanisms of protein import into mitochondria. Curr. Biol. 13, R326–R337.

Tsai, S.L., Oh, J., Singh, S., Chen, R., Chen, W., 2009. Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. Appl. Environ. Microbiol. 75, 6087–6093.

Tyo, K.E., Alper, H.S., Stephanopoulos, G.N., 2007. Expanding the metabolic engineering toolbox: more options to engineer cells. Trends Biotechnol. 25, 132–137.

Varma, A., Palsson, B.O., 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl. Environ. Microbiol. 60, 3724–3731.

Wang, H., Isaacs, F., Carr, P., Sun, Z., Xu, G., Forest, C., Church, G., Programming cells by multiplex genome engineering and accelerated evolution. Nature, 2009.

Win, M.N., Smolke, C.D., 2008. Higher-order cellular information processing with synthetic RNA devices. Science 322, 456–460.

Wintermute, E.H., Silver, P.A., 2010a. Dynamics in the mixed microbial concourse. Genes Dev. 24, 2603–2614.

Wintermute, E.H., Silver, P.A., 2010b. Emergent cooperation in microbial metabolism. Mol. Syst. Biol. 6, 407.

Yanes, O., Tautenhahn, R., Patti, G.J., Siuzdak, G., Expanding Coverage of the metabolome for global metabolite profiling. Anal. Chem., 2011.

Yeates, T.O., Thompson, M.C., Bobik, T.A., The protein shells of bacterial microcompartment organelles. Curr. Opin. Struct. Biol., 2011.

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., Shlomi, T., 2010. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics 26, i255–i260.

Young, I.M., Crawford, J.W., 2004. Interactions and self-organization in the soil-microbe complex. Science (New York, NY) 304, 1634–1637.

Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., Alon, U., 2004. Just-in-time transcription program in metabolic pathways. Nat. Genet. 36, 486–491.

Zhou, Z.H., McCarthy, D.B., O'Connor, C.M., Reed, L.J., Stoops, J.K., 2001. The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. Proc. Natl. Acad. Sci. USA 98, 14802–14807.

Please cite this article as: Boyle, P.M., Silver, P.A., Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. (2011), doi:10.1016/j.ymben.2011.10.003

Appendix D

A New Approach to an Old Problem: Synthetic Biology Tools for Human Disease and Metabolism[1]

[1] Originally published as:
Burrill, D. R.,* Boyle, P. M.* & Silver, P. A. A New Approach to an Old Problem: Synthetic Biology Tools for Human Disease and Metabolism. *Cold Spring Harb Symp Quant Biol* (2011).doi:10.1101/sqb.2011.76.010686. *Equal contribution. Reprinted with permission from the editor.

A New Approach to an Old Problem: Synthetic Biology Tools for Human Disease and Metabolism

D.R. Burrill[1,3], P.M. Boyle[1,3], and P.A. Silver[1,2]
[1]*Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115*
[2]*The Wyss Institute of Biologically Inspired Engineering, Boston, Massachusetts 02115*
Correspondence: pamela_silver@hms.harvard.edu

The field of synthetic biology seeks to develop engineering principles for biological systems. Along these lines, synthetic biology can address human metabolic disease through the development of genetic approaches to the study and modification of metabolism. The re-engineering of natural metabolic states provides fundamental understanding of the integrated components underlying dysfunctional metabolism. Alternatively, the development of biological devices that can both sense and affect metabolic states could render unique control over disease states. In this chapter, we discuss the advancement of synthetic biological approaches to monitoring and engineering metabolism, as well as prospects for synthetic biology's future role in the prevention and treatment of metabolic disease.

One goal of synthetic biology is to develop engineering principles and tools for biology. The characterized components of fundamental biological processes provide a platform for uniquely redesigned biological parts and devices, and the processes themselves can be re-engineered for improved or alternative functions. Since the advent of recombinant DNA technology in the 1970s, genes, promoters, and other biological parts have become modularized and rearranged in increasingly complex manners. During the last decade, the development of cost-effective DNA sequencing and de novo synthesis has accelerated the generation of new biological parts and bioengineering techniques (Boyle and Silver 2009). As our ability to engineer biological systems increases, so too does our ability to address existing biological problems.

The study and treatment of human disease has benefitted from recombinant DNA and synthetic biology approaches. Disease initiation and progression are strongly linked to impaired pathways of metabolic, regulatory, or protein complexes (Li et al. 2011). Complex dysfunctions that are not adequately addressed by small-molecule drugs or peptides may be repaired with synthetic biological devices. Correcting aberrant metabolism is arguably the most challenging pathway to address, given the difficulty in monitoring intracellular metabolites. Biological devices that sense and respond to metabolism may be better equipped to correct metabolism at the cellular scale. Within this chapter, the terms *dysfunctional metabolism* and *metabolic disease* generally describe any disorder characterized by metabolic dysregulation. This includes diabetes, obesity, cancer, cardiovascular disease, Alzheimer disease, and Crohn disease (Suhre et al. 2011).

Whereas some of these metabolic diseases are due to the altered or missing function of a single gene, many are complex polygenic disorders that have proven difficult to cure (Newgard 1992). The initial promise of recombinant DNA technology to cure genetic disorders has not been completely realized (Newgard 1992). It was quickly discovered that greater understanding of existing metabolic pathways and their dysfunction is fundamental to improving disease treatment, including diseases stemming from a single affected gene. The accumulation of metabolic pathway knowledge and model-based evaluation of dysfunction has facilitated our ability to engineer natural metabolism. These advancements, combined with improved gene synthesis and gene-transfer technologies, have also permitted the construction of synthetic devices that can sense and affect metabolism (for review, see Weber and Fussenegger 2009). In this chapter, we discuss both the engineering of natural metabolism and the construction of synthetic devices for metabolism and show how these two approaches have established synthetic biology as a valuable tool against metabolic disease.

ENGINEERING HUMAN METABOLISM

Metabolic engineering—the application of recombinant DNA technology to direct carbon flux toward production of a desired compound—has driven our understanding of metabolic control in recent years. Although focused primarily on microbial cell engineering, this research has revealed principles that are broadly applicable to human metabolism. Furthermore, synthetic tools for manipulating metabolic flux can be used to correct disease-associated metabolic defects. Therapeutics derived through synthetic biology for metabolic diseases are likely to build upon these engineering efforts in simpler systems.

[3]These authors contributed equally to this work.

 doi: 10.1101/sqb.2011.76.010686

A challenge in metabolic engineering is that control of metabolic flux is often distributed throughout a pathway, rather than a single enzyme serving as a *rate-limiting step* (Thomas and Fell 1996; Fell 1997). Consequently, engineered pathways must typically be *tuned* to maximize throughput, often by adjustment of enzyme concentrations (Dekel and Alon 2005; Boyle and Silver 2009; Wang et al. 2009). Multisite control is likely to be a property of biological devices that reroute metabolic flux in the human body.

To achieve quantitative control over metabolism, a wide array of biological control elements, such as promoters, ribosome-binding sites, and riboswitches, have been repurposed or synthesized (Boyle and Silver 2009). The efficacy and robustness of control elements are especially critical in a therapeutic context, as compared with traditional metabolic engineering. Microbes are often engineered to simply act as *pipes* for chemical conversions, tuned to continuously produce a target compound. In contrast, biological devices destined for the human body will require stringent logic and control. For example, bacteria that target tumors for destruction must distinguish between normal and cancerous tissue, to avoid destroying healthy cells (Anderson et al. 2007).

Achieving an appropriately strict regulation of metabolic networks requires a quantitative understanding of network properties. Models of microbial metabolism have been instrumental to many metabolic engineering efforts (Burgard et al. 2003; Pharkya et al. 2004; Bro et al. 2006; Kennedy et al. 2009; Wintermute and Silver 2010). In the context of human biology, models are important tools in analyzing metabolomic data, predicting drug behavior, and identifying metabolic flux modes that characterize disease. Because the assistive tools of mutagenesis and selection are not available for human systems, models are vital to understanding human metabolism.

Modeling frameworks such as Flux Balance Analysis (FBA) and its derivatives have been instrumental in the engineering of microbial metabolism (Varma and Palsson 1994; Boyle and Silver 2009; Gianchandani et al. 2010). FBA-type models take advantage of our near-comprehensive understanding of the stoichiometry of metabolic networks across many species. In essence, these models incorporate all known metabolic reactions in an organism and identify a steady-state flux distribution that maximizes or minimizes a user-defined *objective function* (Feist and Palsson 2010). To increase their biologic relevance and minimize the number of optimal flux distributions, FBA-type models further incorporate constraints such as limits on nutrient-uptake rate. The details on the formulation, selection of constraints and objective functions, optimization strategies, and practical applications of these models have been extensively reviewed (Fig. 1) (Varma and Palsson 1994; Edwards et al. 2002; Oberhardt et al. 2009; Terzer et al. 2009). In this chapter, we present progress in the adaptation of these models to mammalian systems.

Mammalian biology is inherently more complex than that of bacteria or yeast. Models of mammalian tissues, tumors, or cell lines must integrate constraints on gene expression and multicellular interactions that are representative of a given tissue type. These parameters are often inferred from microarray datasets, starting with a generic model of human metabolism and removing unexpressed enzymes (Duarte et al. 2007; Jerby et al. 2010). Cell-specific extracellular metabolite availability and objective functions improve model accuracy. The relatively low-parameter requirements of FBA-type models also allows multicellular metabolic interactions to be modeled (Stolyar et al. 2007; Wintermute and Silver 2010).

Most problematic for constraint-based approaches is the formulation of appropriate objective functions for mammalian cells. When modeling microbes, it is commonly assumed that laboratory-adapted strains such as *Escherichia coli* are tuned for rapid growth (Ibarra et al. 2002). In contrast, many cells in the human body are not constantly dividing (Berk 2005). Even proliferative cancer cells are likely to be governed by more complex objectives than rapid growth. Rather than defining an objective function a priori, many researchers have instead opted to infer the wild-type flux distribution from transcriptomic and proteomic data (Shlomi et al. 2008; Jerby et al. 2010; Lewis et al. 2010).

A multitissue model of brain metabolism was recently constructed to study Alzheimer disease (Fig. 1) (Lewis et al. 2010). Multiple "omics" datasets guided construction of metabolic networks relevant to each modeled cell type: astrocytes, glutamatergic neurons, GABAergic neurons, and cholinergic neurons. Relevant extracellular metabolic reactions such as those related to the endothelium, blood supply, and the interstitial spaces between cells were also included.

Because gene deletion simulations are a common application of constraint-based approaches, these models of neuron metabolism were used to simulate enzyme deficiencies common to Alzheimer disease, such as decreased mitochondrial pyruvate dehydrogenase activity (Bubber et al. 2005; Lewis et al. 2010). This simulation is achieved by removing a network reaction and recalculating the flux distribution of the mutant network (Segrè et al. 2002; Burgard et al. 2003; Duarte et al. 2004; Kennedy et al. 2009; Rocha et al. 2010). In lieu of an objective function for brain energy metabolism, researchers instead used Monte Carlo sampling to identify feasible flux distributions (Lewis et al. 2010). These simulations provided mechanistic insight into the expression changes observed in microarray datasets, including the identification of neuronal pathways that may limit synthesis of acetyl-CoA and, by extension, the neurotransmitter acetylcholine in Alzheimer neurons.

Network-scale models of metabolism may be instrumental in the analysis of complex metabolic dysfunction. For example, cancer genome studies have revealed that cancer cells often contain multiple gene-copy-number changes and mutations (Beroukhim et al. 2010). Evidence of multisite control in cancer metabolism suggests that several enzymes must be modulated to correct or halt metabolism in cancer cells (Rasnick and Duesberg 1999; Moreno-Sánchez et al. 2010; Schrattenholz et al. 2010). Metabolic models that can integrate omics data will be

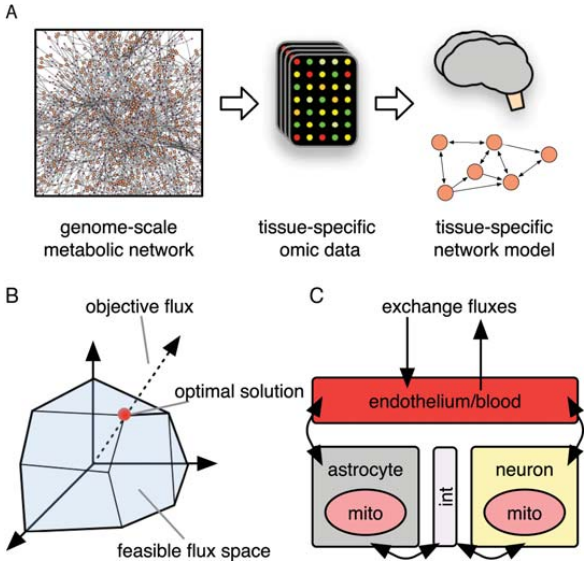


Figure 1. Constraint-based modeling of mammalian tissues. (*A*) Tissue-specific constraint-based models start with a genome-scale metabolic network. These networks, such as *Homo sapiens build 1* (Duarte et al. 2007; http://gcrg.ucsd.edu/Organisms), contain all possible metabolic reactions defined in the human genome. Tissue-specific datasets, such as expression microarrays, are then used to define the particular subset of reactions that occurs in a given tissue. This reduced network is then used to model the tissue of interest. (*B*) A visualization of constraint-based modeling. Constraints on metabolic fluxes, input rates, and output rates define the bounds of the feasible flux space for a given metabolic network. To determine which flux mode is most likely to exist in the tissue to be modeled, an objective function is chosen, which defines the metabolic flux that the system seeks to optimize. Although many solution algorithms exist, most involve the intersection of the objective function with the surface of the feasible flux space. (*C*) Compartmentalization of a human brain metabolic model (Lewis et al. 2010). Cell-specific models were defined for astrocytes, glutamatergic neurons, GABAergic neurons, and cholinergic neurons. Metabolic networks for the endothelium, blood, and interstitial space (int) connected the cell-specific metabolic networks. Mitochondria (mito) for each cell type were also modeled as discrete metabolic compartments within each cell type. Exchange fluxes, the input and output rates of metabolites, were experimentally measured and used to fine-tune the model.

useful in diagnosing and identifying therapeutic strategies. As in the Alzheimer disease example, inferring metabolic objective functions of cancer cells will identify metabolic pathways to be targeted for modulation with drugs or synthetic devices.

SYNTHETIC DEVICES FOR METABOLIC DYSFUNCTION

Given a quantitative understanding of metabolic dysfunction, we must then be able to correct these dysfunctions in a predictable manner. Engineered biological pathways termed *synthetic circuits* or *devices* may be designed to execute useful functions within single living human cells. Synthetic circuits offer a more controlled approach to traditional drug and gene therapies, such as the ability to dynamically silence, activate, and tune the expression of desired genes and drugs (Khalil and Collins 2010). Simply put, synthetic devices detect an environmental signal (e.g., hypoxia) that is processed by a genetic logic circuit, which filters the signal and activates a cellular response (e.g., drug delivery) (Fig. 2). Such circuits can communicate a cell's metabolic state and provide a therapeutic biological action. Organisms ranging from bacteria to mammalian cells provide an abundance of existing modular sensory regulatory circuits, the components of which can be transferred to synthetic circuits in human cells. Using cues from nature, synthetic circuits have been designed to regulate cells at the transcriptional, translational, or posttranslational level for the purpose of studying and treating metabolic disease.

TRANSCRIPTIONAL REGULATION

One way that cells mobilize a response to metabolic perturbations is via transcription. In simple terms, an input signal modulates the binding of transcription factor proteins at binding sites near promoters, conferring a gene activity state. Following this design, synthetic circuits have been engineered with endogenous or artificial transcription factors fused to activators or repressors that bind natural or synthetic elements within a minimal promoter to control gene output. Sensitivity and specificity can be introduced to this synthetic system by various methods (for review, Burrill and Silver 2010): adding, subtracting, or mutating binding sites; using activators and repressors of varying strengths; or inserting intronic sequences in promoter regions to tune transcriptional

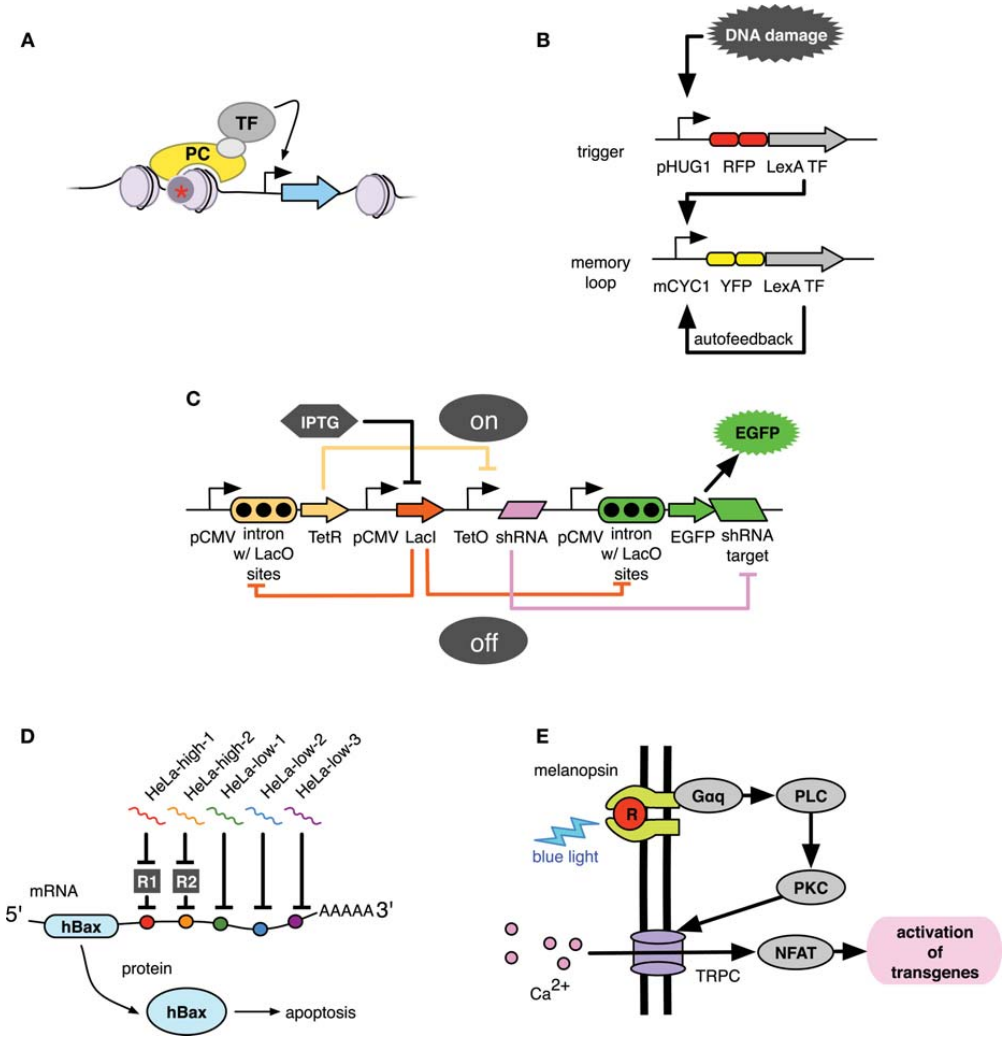


Figure 2. Synthetic devices for metabolic dysfunction. (*A*) *Synthetic reversal of epigenetic silencing* (Haynes and Silver 2011). A Polycomb chromodomain (PC) fused to a transcription factor (TF) binds the methylation mark H3K27me3 (star) on chromatin (circles), thereby activating gene expression. (*B*) *Synthetic memory device* (Burrill and Silver 2011). DNA damage transiently activates a red fluorescent protein (RFP)-labeled *trigger* gene via a damage-inducible promoter from the yeast gene *HUG1*. The trigger gene produces the transcription factor LexA fused to a DNA-binding domain. The trigger transactivator subsequently induces *memory loop* gene expression, producing a yellow fluorescent protein (YFP)-labeled LexA transcription factor that binds to corresponding sites located within its own promoter. After the DNA damaging agent is removed, the memory loop remains stably expressed. (*C*) *Synthetic switch* (Deans et al. 2007). In the *on* state, IPTG binds to LacI repressor proteins (orange), thus freeing *lac* operator (LacO) sites and permitting EGFP and TetR transcription. TetR proteins (yellow) further enhance EGFP expression by binding to a *tet* operator (TetO) site located upstream of the RNAi module (pink), thus repressing short hairpin RNA (shRNA) transcription. In the *off* state, LacI repressor proteins are constitutively expressed and bind to LacO sites in the transgene, thereby repressing EGFP expression. LacI repressor proteins also bind to LacO sites in TetR, causing TetR repression. Thus, shRNA is transcribed and silences EGFP. (*D*) *Synthetic cell-type detector* (Xie et al. 2011). Using markers characteristic of HeLa cancer cells, a sensor motif for markers expressed high and low was engineered. The HeLa-high motif is comprised of a double-inversion module (R1 and R2) that allows expression only if the marker is present at or above its level in HeLa but represses the output if the marker is low. A HeLa-low marker sensor was implemented by fusing repeats of complementary target sites into the output's 3′-UTR. Multiple sensors were combined by fusing their corresponding microRNA (miRNA) targets in the output's 3′-UTR. If expression levels match that of HeLa cells, the proapoptotic protein hBax is activated. (*E*) *Synthetic signaling cascade* (Ye et al. 2011). Blue-light-mediated photoisomerization of the retinal chromophore triggers melanopsin conformational changes that activate the Gaq-type G protein (Gaq), phospholipase C (PLC), and phosphokinase C (PKC). Consequently, calcium ion influx occurs via transient receptor potential channels (TRPCs), ultimately resulting in activation of the transcription factor, nuclear factor of activated T cells (NFAT), and induction of NFAT-inducible transgenes.

rates (Swinburne et al. 2008). The primary motivation in the design of these control elements is to ensure that a synthetic circuit behaves in a predictable, controlled manner, which is critical when translating such devices to human studies.

Several systems have been designed in which a signal-responsive transcription factor is fused to a DNA-binding domain that binds to an engineered promoter upstream of a reporter gene. This design has been used to report on hormone concentrations (Braselmann et al. 1993), hypoxia (Tang et al. 2005), redox levels (Weber et al. 2006), and chromatin modifications (Hansen et al. 2008). In recent work by Haynes et al. (Haynes and Silver 2011), Polycomb chromatin protein was used to engineer artificial transcription factors that recognize repressive histone methylation marks and activate silenced, disease-related genes in human cancer cells (Fig. 2A). In other examples, custom-built transcription factors composed of zinc-finger DNA-binding domains respond to external stimuli by targeting specific mammalian genes. Endogenous genes that are silenced in many cancers—such as the anti-angiogenic *pigment epithelium derived factor* (Yokoi et al. 2007) and the breast tumor suppressor *maspin* (Beltran et al. 2007)—have been reactivated in mammalian cell culture using this method.

Other transcription-based biosensors employ minimal promoters with natural regulatory DNA elements that are activated by environment-responsive, endogenous transcription factors, producing sensors capable of reporting the presence of progesterone (Strähle et al. 1987), DNA damage (Ohno et al. 2008), and glucocorticoid hormone (Meijsing et al. 2009). Circuits such as those described are capable of providing valuable information about the internal metabolic environment. This information can then be potentially used for restoring misregulated metabolism in a therapeutic setting. Furthermore, the described devices are highly modular, making them amenable to the study of multiple diseases.

In some cases, it might be desirable to use a synthetic circuit that controls transcription by processing more complex input logic. Examples include memory circuits and oscillators, both of which naturally mediate gene expression to control metabolism, signaling pathways, and cell differentiation. Biological memory can be defined as a sustained cellular response to a transient stimulus (Burrill and Silver 2011). One way that cells accomplish this task is through bistable transcriptional states, such that a chemical state becomes defined as *on* or *off* in response to environmental input and, given certain parameters, can be inherited through DNA replication and cell division. Synthetic memory circuits modeled from this design are capable of sensing external inputs and producing permanent expression of a reporter (e.g., fluorescent protein, specific gene, or drug). Two have been recently designed in yeast—both use transcriptional positive feedback to convert a transient exposure to galactose (Ajo-Franklin et al. 2007) or DNA damage (Burrill and Silver 2011) into sustained expression of a fluorescent reporter (Fig. 2B). If transferred to mammalian systems, memory devices such as these could be used to study the long-term effects of particular stimuli on a cell population or provide long-term gene or drug dosage in response to transient inputs.

Synthetic mammalian oscillators based on their natural counterparts can control the periodic expression of desired genes for therapeutic reasons. Indeed, most hormones are released in time-dependent pulses, and replicating these patterns in hormone dosage to patients may decrease side effects and improve response (Khalil and Collins 2010). One circuit that could be used for such a purpose is the sense–antisense transcriptional device engineered by Tigges et al. (2009). The device encodes a positive and a time-delayed negative-feedback loop that enables self-sustained oscillatory mammalian gene expression. The work shows that oscillation frequency, amplitude, and damping rates can be tuned by adding regulatory elements (e.g., repressors, activators) responsive to artificial chemical inputs or by altering the network topology. Another engineered oscillator allows for pulses of gene expression via negative feedback. The pulses can vary in transcription time by using increasingly long introns to lengthen the time required for gene transcription (Swinburne et al. 2008). Such a device could be used to tightly regulate synthetic gene expression in mammalian systems.

A final example of a more complex transcription regulator is a synthetic genetic counter (Friedland et al. 2009). Control over drug release might be integrated by programming a synthetic system to self-destruct after a specific number of gene or drug release pulses or cell cycles. Many of these transcriptional tools to control drug and gene delivery have yet to be successfully engineered in human cells, but their development in bacteria and yeast is laying the groundwork for future translational applications.

TRANSLATIONAL REGULATION

RNA molecules are an attractive target for bioengineering owing to their many functional capacities. These functions include splicing RNA, catalyzing biochemical reactions, and regulating gene expression (Khalil and Collins 2010). Synthetic biologists have used RNA to achieve direct and tight regulation of protein concentrations primarily via RNA interference (RNAi) and functional RNAs (ribozymes, riboswitches) (Haynes and Silver 2009).

RNAi uses sequence complementarity to silence target messenger RNA (mRNA). As such, RNAi constructs can be designed against any desired RNA. Bioengineers have used RNAi to tighten control and tweak the expression dynamics of synthetic systems. Greber et al. (2008) demonstrated that intronic short interfering RNAs (siRNAs) can reduce basal gene expression of inducible transcription systems (Greber et al. 2008). Deans et al. (2007) designed an inducible switch in mammalian cells that produces shRNAs to prevent leaky expression of transgenes (Fig. 2C).

Other synthetic devices use functional RNAs to control circuit behavior (Bayer and Smolke 2005; Beisel et al.

2008). Ribozymes, antiswitches, and shRNA switches contain an aptamer domain (sensor) involved in specific ligand recognition and an expression platform that controls gene activity by causing mRNA conformational changes (Mulhbacher et al. 2010). The flexibility and ease of design of synthetic aptamer domains for a wide range of target ligands makes them a valuable synthetic biology tool (Win and Smolke 2009). Ribozymes can specifically cleave target RNA and have been effectively used in gene therapy for cancer and other diseases (Win and Smolke 2008; Culler et al. 2010; Mulhbacher et al. 2010). For example, a hammerhead ribozyme targeting the human epidermal growth factor receptor HER-2 reversed the malignant phenotype of breast cancer cells (He et al. 2010). Metabolite-binding riboswitches also represent a novel solution to sensing and studying metabolic disease. To date, synthetic riboswitches have been built to sense thiamine pyrophosphate, lysine, flavin mononucleotide, and guanine (Mulhbacher et al. 2010). The pairing of switchable aptamer domains with ribozymes that degrade an mRNA in a ligand-dependent manner allows a rapid regulatory response to changes in intracellular metabolite concentrations (Win and Smolke 2008). The construction of aptamer domains responsive to other environmental conditions or metabolites could serve as future therapeutic and research tools for disease.

Other RNAi-based synthetic devices can process multiple inputs, making it possible to specifically detect diseases characterized by several metabolic signals. A mammalian logic evaluator composed of genes containing multiple synthetic RNAi target sequences was shown to correctly detect up to five siRNA inputs (Rinaudo et al. 2007). In a recent paper, Xie et al. (2011) showed that expression of endogeneous miRNAs characteristic of cervical cancer cells could trigger apoptosis via a synthetic transcriptional device regulating expression of human Bcl-2-associated X protein hBax (Fig. 2D). Circuits such as these are useful in situations where a single metabolic signal does not define an environment. The continued engineering of logic circuits that evaluate endogenous cellular inputs will lead the way to synthetic devices that can sense and correctly respond to complex physiological conditions.

CELL-SIGNALING REGULATION

At the posttranslational level, signal transduction pathways represent another platform for synthetic regulation of human cells. Cell-signaling pathways are often misregulated in metabolic diseases and are thus a target for therapeutic intervention. The sensor element for many signaling pathways is a membrane-bound or nuclear protein receptor. Whereas environment-responsive promoters and RNA aptamers are derived from nature or selections, protein receptors can be designed at the level of molecular interaction (Khalil and Collins 2010). Existing signaling pathways provide many further components that can be harnessed to build tunable, sensitive, and specific synthetic pathways, such as the number and spacing of proteins in a given cascade. Thus, signal transduction pathways constitute a modular method for programming protein-based regulators.

The construction of chimeric protein ligands that bind receptors in a precise and predictable manner can increase the specificity of synthetic devices for target cell types. For example, a chimera was designed to specifically bind to cells with both epidermal growth factor (EGF) receptor and interferonα-2a (IFNα-2a) receptor (Cironi et al. 2008). These authors demonstrated that a series of IFNα-2a mutations could progressively decrease the on rate and dissociation constant of the IFNα-2a-IFNα receptor 2 (IFNAR2) interaction (Cironi et al. 2008). This quantitative knowledge was then applied to the design of chimeras in which IFNα receptor activation depends on the presence of EGF receptor on the same cell. This work illustrates the value of a quantitative approach to tool design and suggests useful techniques for modulating protein–protein interactions.

A number of synthetic approaches have addressed the abnormal cell signaling associated with diabetes mellitus. Diabetes develops when insulin is inadequately produced by the pancreas or not effectively used by cells, resulting in the failure of glucose absorption from blood for cellular energy. It has been argued that cell-based treatments for insulin-dependent diabetes may provide more physiologic regulation of blood glucose levels than daily insulin injections (Zhang et al. 2008). Surrogate β cells have been developed to synthesize functional insulin in response to changing glucose levels, thus re-establishing a normal metabolic signaling pathway. This was accomplished both in human cell culture (Bara and Sambanis 2008) and mouse models (Zhang et al. 2008), the latter revealing that blood glucose levels could be reduced to normal using this method. These studies suggest that the use of genetically engineered cells to express human insulin might be a powerful therapeutic approach to diabetes treatment.

Other synthetic approaches to diabetic regulation have involved inserting entirely new signaling pathways into human cells and mice. In one example, the investigators designed a light-inducible signaling cascade to control transgene expression of the glucagon-like peptide 1, a protein that increases insulin secretion. The implantation of the device in diabetic mice coincided with the dissipation of glycemic excursions (Fig. 2E) (Ye et al. 2011). In another system, the glyoxylate shunt pathway was introduced to mammalian liver cells and mice (Dean et al. 2009). Mice expressing the shunt exhibited increased fatty-acid oxidation and decreased tendencies toward diet-induced obesity. This work offers new synthetic tools for studying and manipulating biological networks to correct metabolic disorders.

TOOL DELIVERY

Decades of experience in molecular genetics and gene therapy have demonstrated the importance of caution when designing novel genetic therapeutics (Denèfle 2011). In developing therapeutic synthetic devices for human cells, it is important to anticipate deleterious off-

target effects and host–cell interactions. In cases where engineered human cells need to be delivered to the human body, the cells will probably need to be implanted (Ausländer et al. 2011). To prevent immune system rejection of nonautologous cells and the side effects of immunosuppressive drugs, one might want to encapsulate engineered cells in a biocompatible, semipermeable membrane. This technology has been limited so far by inadequate nutrient transport and undesirable immune responses (Ausländer et al. 2011). Recent efforts have focused on developing submicron scale coatings to reduce cell mass and incorporating anti-inflammatory capabilities into synthetic membranes (Ausländer et al. 2011). This area of work will have to carefully but continually improve to actualize the potential of cell-based therapeutics.

Delivery methods and efficacy will also depend on the cell type to be engineered. Immortalized cancer cell lines common to laboratory research are unlikely to be utilized in human therapies. Engineered stem cells, however, are more stable and predictable to engineer and have already been used safely in several clinical trials (Denèfle 2011; Trounson et al. 2011). In particular, mesenchymal stem cells (MSCs) possess many advantageous cellular characteristics, such as ease of isolation, high expansion potential, and genetic stability. For these reasons, MSCs are popular therapeutic and engineering platform candidates (Le Blanc and Ringdén 2007).

An alternative to engineering human cells is to design synthetic microbial devices that interact with human cells. Several factors make the prospect of synthetic microbial therapeutics more likely in the near term than devices that modify mammalian cells. Microbial engineering is generally easier because microbes have smaller genomes, grow faster, and are more genetically accessible than mammalian cells (Weber and Fussenegger 2009). Furthermore, microbial treatments are likely to be more transient than gene-therapy-type approaches, rendering them potentially safer. Whereas the mislocalization of microbes in the human body can itself cause disease, the human body is home to trillions of symbiotic microbes; engineering these species in their respective niches may be a successful strategy.

For example, synthetic microbial devices might be used in the selective destruction of cancer cells, therein providing the benefits of traditional chemotherapy without the negative side effects (Anderson et al. 2006, 2007; Cironi et al. 2008). To this end, cancer-targeting bacteria have been developed in which their ability to invade and kill cancer cells is linked to environmental signals that are specific to the tumor microenvironment (such as hypoxia and bacterial aggregation). At the center of this device is a modular AND gate that triggers cell invasion only if the two inputs are detected (Anderson et al. 2006). This system is composed of modular parts, such that the system is adaptable to different cancer types. Recently, nonpathogenic or immunogenic photosynthetic bacteria were engineered with invasin and listeriolysin to invade the cytoplasm of mammalian cells (Agapakis et al. 2011). This work presents the possibility of using photosynthetic bacteria as a novel chassis for synthetic devices in human systems.

It should be noted that the safe delivery of both mentioned bacterial devices to patients may be problematic, because an immune response against engineered bacteria would be highly undesirable. The human gut is a likely target for engineered bacteria. The average person harbors approximately 160 bacterial species in his or her gut, with about 10^{14} individual microbial cells (Qin et al. 2010; Tilg and Kaser 2011). The variety of species found in a single person can depend on the person's location, diet, and disease state, among other factors. Cross talk between the host and the microbiome is also implicated in disease risk and progression, including obesity, cardiovascular disease, diabetes, cancer, Crohn's disease, ulcerative colitis, and the metabolic syndrome (Martins Dos Santos et al. 2010; Tilg and Kaser 2011). There is tremendous potential for engineered microbes to observe, diagnose, and potentially treat these diseases.

The species composition (and therefore the metabolic state) of the gut microbiome is maintained through intercellular communication, both between the host and the microbiome and within the microbiome itself. Ideally, engineered microbes in the gut will participate in this communication to maintain homeostasis. Two signaling peptides, glucagon-like peptide 1 (GLP-1) and pancreatic and duodenal homeobox gene 1 (PDX-1), have been demonstrated to induce insulin synthesis in the intestinal epithelium (Yoshida et al. 2002; Suzuki et al. 2003). *E. coli* engineered to secrete GLP-1 and PDX-1 in response to glucose were demonstrated to trigger insulin production in Caco-2 cells (Duan et al. 2008). Introduction of this device into the gut of diabetic patients could restore glucose-dependent insulin production. A similar device, this time secreting *Vibrio cholerae* quorum-sensing molecules, was capable of preventing *V. cholerae* propagation in mice (Duan and March 2010). Prophylactic engineered bacteria have also shown promise in preventing HIV infection of the cervicovaginal mucosa (Rao et al. 2005; Liu et al. 2006; Yu et al. 2009; Duan and March 2010; Lagenaur et al. 2011).

Fecal transplantation is emerging as a clinical treatment for gastrointestinal infections, particularly for *Clostridium dificile* (Palmer 2011; Tilg and Kaser 2011). The primary benefit of fecal transplantation is the recolonization of the gastrointestinal tract by commensal microbes that were eradicated during the treatment of the initial infection. The introduction of engineered bacteria to the gastrointestinal tract may prevent infection altogether. Beyond the use of natural probiotic bacteria, or the constitutive secretion of recombinant peptides by engineered bacteria, we envision that bacteria equipped with synthetic gene circuits might sense unwanted changes in the body and react by secreting drugs or signaling molecules. This secondary immune system could augment the body's natural defenses to fight disease at preclinical stages.

CONCLUSIONS AND OUTLOOK

In this chapter, we have provided an overview of the synthetic biology tools available for human disease and metabolism. Whereas some of these techniques have been

successfully applied to human patients, many synthetic biology tools remain at the laboratory bench. Successful transition to the clinic will require the development of robust delivery methods and significant improvement in device reliability and specificity. These objectives are in line with the ultimate goals of synthetic biology—to build biological systems with defined and predictable functions. Ultimately, *smart* synthetic devices, which target themselves to the correct tissues and regulate their own dosage, may be safer and more effective than current therapeutics.

We have touched upon existing and proposed methods for introducing synthetic tools to the human body. To date, synthetic biologists have constructed a diverse array of components that may contribute to future therapeutics, but assembly of a complete device that is ready for human testing has yet to be realized. Therapeutic implantation of synthetic biological devices is but one application of these tools; synthetic biological devices have already enhanced our understanding of disease and metabolism.

The maturation of synthetic biology in the medical arena is dependent on further scientific and regulatory development. To be considered for clinical trials, the reliability and accuracy of synthetic devices must be increased. The limited number of animal experiments conducted thus far (Duan and March 2010; Lagenaur et al. 2011; Ye et al. 2011) speaks to the infancy of the field. The complexity of engineered mammalian devices has tended to lag behind progress in the synthetic biology of simpler systems. The modularization of genetic circuits, that is, repurposing and combining devices to produce new functions or operate in different cell types, has been difficult. Our own experience with memory devices (and the experience of others with logic gates) has taught us that devices can be modularized, but that more than a simple promoter swap is often necessary (Ajo-Franklin et al. 2007; Anderson et al. 2007; Burrill and Silver 2011). These issues are not insurmountable, but they do lengthen the time horizon for building clinical-grade devices. Navigating the complex pharmaceutical regulatory landscape will present an entirely different yet arduous challenge. These issues raise the importance of accelerating the mammalian engineering design cycle: We must expand work in therapeutically relevant cell lines, incorporate improved models of device function, and better integrate with the related fields of stem cell therapy and tissue engineering.

The stakes for human applications of synthetic biology are higher than for applications confined to bioreactors. In developing synthetic biology tools for human health, we must continually consider the relevant ethical and social issues surrounding this research (Yearley 2009). Taking initiative in these areas will be essential to the responsible development of synthetic biological therapeutics. Recent experience in the development of experimental therapies has underscored the importance of caution in developing human trials (Marshall 1999; Gabardi et al. 2011). Balancing risks with the transformative potential benefits of synthetic biology will be a continual and necessary challenge as synthetic biology enters the clinic.

ACKNOWLEDGMENTS

We sincerely thank Mara Inniss for helpful discussion and a critical reading of the manuscript. D.R.B. is supported by the National Science Foundation (NSF) Synthetic Biology Engineering Research Center. P.M.B. is supported by the Harvard University Center for the Environment and the NSF Synthetic Biology Engineering Research Center. P.A.S. is supported by the National Institutes of Health (grants GM36373-22 and GM36373).

REFERENCES

Agapakis CM, Niederholtmeyer H, Noche RR, Lieberman TD, Megason SG, Way JC, Silver PA. 2011. Towards a synthetic chloroplast. *PLoS ONE* **6:** e18877.

Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EPS, Landgraf D, Phillips I, Silver PA. 2007. Rational design of memory in eukaryotic cells. *Genes Dev* **21:** 2271–2276.

Anderson JC, Clarke EJ, Arkin AP, Voigt CA. 2006. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* **355:** 619–627.

Anderson JC, Voigt CA, Arkin AP. 2007. Environmental signal integration by a modular AND gate. *Mol Syst Biol* **3:** 133.

Ausländer S, Wieland M, Fussenegger M. 2011. Smart medication through combination of synthetic biology and cell microencapsulation. *Metab Eng* doi: 10.1016/j.ymben.2011.06.003.

Bara H, Sambanis A. 2008. Insulin-secreting L-cells for the treatment of insulin-dependent diabetes. *Biochem Biophys Res Commun* **371:** 39–43.

Bayer TS, Smolke CD. 2005. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23:** 337–343.

Beisel CL, Bayer TS, Hoff KG, Smolke CD. 2008. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol* **4:** 224.

Beltran A, Parikh S, Liu Y, Cuevas BD, Johnson GL, Futscher BW, Blancafort P. 2007. Re-activation of a dormant tumor suppressor gene maspin by designed transcription factors. *Oncogene* **26:** 2791–2798.

Berk AJ. 2005. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. *Oncogene* **24:** 7673–7685.

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, et al. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* **463:** 899–905.

Boyle PM, Silver PA. 2009. Harnessing nature's toolbox: Regulatory elements for synthetic biology. *J R Soc Interface* **6** (Suppl 4)**:** S535–S546.

Braselmann S, Graninger P, Busslinger M. 1993. A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. *Proc Natl Acad Sci* **90:** 1657–1661.

Bro C, Regenberg B, Förster J, Nielsen J. 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng* **8:** 102–111.

Bubber P, Haroutunian V, Fisch G, Blass JP, Gibson GE. 2005. Mitochondrial abnormalities in Alzheimer brain: Mechanistic implications. *Ann Neurol* **57:** 695–703.

Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84:** 647–657.

Burrill DR, Silver PA. 2010. Making cellular memories. *Cell* **140:** 13–18.

Burrill DR, Silver PA. 2011. Synthetic circuit identifies subpopulations with sustained memory of DNA damage. *Genes Dev* **25:** 434–439.

Cironi P, Swinburne IA, Silver PA. 2008. Enhancement of cell type specificity by quantitative modulation of a chimeric ligand. *J Biol Chem* **283:** 8469–8476.
Culler SJ, Hoff KG, Smolke CD. 2010. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **330:** 1251–1255.
Dean JT, Tran L, Beaven S, Tontonoz P, Reue K, Dipple KM, Liao JC. 2009. Resistance to diet-induced obesity in mice with synthetic glyoxylate shunt. *Cell Metab* **9:** 525–536.
Deans TL, Cantor CR, Collins JJ. 2007. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130:** 363–372.
Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436:** 588–592.
Denèfle PP. 2011. Introduction to gene therapy: A clinical aftermath. *Meth Mol Biol* **737:** 27–44.
Duan F, March JC. 2010. Engineered bacterial communication prevents *Vibrio cholerae* virulence in an infant mouse model. *Proc Natl Acad Sci* **107:** 11260–11264.
Duan F, Curtis KL, March JC. 2008. Secretion of insulinotropic proteins by commensal bacteria: Rewiring the gut to treat diabetes. *Appl Environ Microbiol* **74:** 7437–7438.
Duarte NC, Herrgård MJ, Palsson BØ. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **14:** 1298–1309.
Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci* **104:** 1777–1782.
Edwards JS, Covert M, Palsson B. 2002. Metabolic modelling of microbes: The flux-balance approach. *Environ Microbiol* **4:** 133–140.
Feist AM, Palsson BØ. 2010. The biomass objective function. *Curr Opin Microbiol* **13:** 344–349.
Fell D. 1997. *Understanding the Control of Metabolism*, chapters 4 and 5. Portland Press, London.
Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. 2009. Synthetic gene networks that count. *Science* **324:** 1199–1202.
Gabardi S, Halloran PF, Friedewald J. 2011. Managing risk in developing transplant immunosuppressive agents: The new regulatory environment. *Am J Transplant* **11:** 1803–1809.
Gianchandani EP, Chavali AK, Papin JA. 2010. The application of flux balance analysis in systems biology. *Wiley Interdiscip Rev Syst Biol Med* **2:** 372–382.
Greber D, El-Baba MD, Fussenegger M. 2008. Intronically encoded siRNAs improve dynamic range of mammalian gene regulation systems and toggle switch. *Nucl Acids Res* **36:** e101.
Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappsilber J, Lerdrup M, Helin K. 2008. A model for transmission of the H3K27me3 epigenetic mark. *Nature* **10:** 1291–1300.
Haynes KA, Silver PA. 2009. Eukaryotic systems broaden the scope of synthetic biology. *J Cell Biol* **187:** 589–596.
Haynes KA, Silver PA. 2011. Synthetic reversal of epigenetic silencing. *J Biol Chem* **286:** 27176–27182.
He P, Zhu D, Hu J-J, Peng J, Chen L-S, Lu G-X. 2010. pcDNA3.1(−)-mediated ribozyme targeting of HER-2 suppresses breast cancer tumor growth. *Mol Biol Rep* **37:** 1597–1604.
Ibarra RU, Edwards JS, Palsson BØ. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420:** 186–189.
Jerby L, Shlomi T, Ruppin E. 2010. Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Mol Syst Biol* **6:** 401.
Kennedy CJ, Boyle PM, Waks Z, Silver PA. 2009. Systems-level engineering of nonfermentative metabolism in yeast. *Genetics* **183:** 385–397.
Khalil AS, Collins JJ. 2010. Synthetic biology: Applications come of age. *Nat Rev Genet* **11:** 367–379.
Lagenaur LA, Sanders-Beer BE, Brichacek B, Pal R, Liu X, Liu Y, Yu R, Venzon D, Lee PP, Hamer DH. 2011. Prevention of vaginal SHIV transmission in macaques by a live recombinant Lactobacillus. *Mucosal Immunol* **4:** 648–657.
Le Blanc K, Ringdén O. 2007. Immunomodulation by mesenchymal stem cells and clinical experience. *J Intern Med* **262:** 509–525.
Lewis NE, Schramm G, Bordbar A, Schellenberger J, Andersen MP, Cheng JK, Patel N, Yee A, Lewis RA, Eils R, et al. 2010. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol* **28:** 1279–1285.
Li X, Li C, Shang D, Li J, Han J, Miao Y, Wang Y, Wang Q, Li W, Wu C, et al. 2011. The implications of relationships between human diseases and metabolic subpathways. *PLoS ONE* **6:** e21131.
Liu X, Lagenaur LA, Simpson DA, Essenmacher KP, Frazier-Parker CL, Liu Y, Tsai D, Rao SS, Hamer DH, Parks TP, et al. 2006. Engineered vaginal lactobacillus strain for mucosal delivery of the human immunodeficiency virus inhibitor cyanovirin-N. *Antimicrob Agents Chemother* **50:** 3250–3259.
Marshall E. 1999. Clinical Trials: Gene therapy death prompts review of adenovirus vector. *Science* **286:** 2244–2245.
Martins Dos Santos V, Müller M, de Vos WM. 2010. Systems biology of the gut: The interplay of food, microbiota and host at the mucosal interface. *Curr Opin Biotechnol* **21:** 539–550.
Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, Yamamoto KR. 2009. DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324:** 407–410.
Moreno-Sánchez R, Saavedra E, Rodríguez-Enríquez S, Gallardo-Pérez JC, Quezada H, Westerhoff HV. 2010. Metabolic control analysis indicates a change of strategy in the treatment of cancer. *Mitochondrion* **10:** 626–639.
Mulhbacher J, Brouillette E, Allard M, Fortier L-C, Malouin F, Lafontaine DA. 2010. Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog* **6:** e1000865.
Newgard CB. 1992. Cellular engineering for the treatment of metabolic disorders: Prospects for therapy in diabetes. *Biotechnology (NY)* **10:** 1112–1120.
Oberhardt MA, Palsson BØ, Papin JA. 2009. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* **5:** 320.
Ohno K, Ishihata K, Tanaka-Azuma Y, Yamada T. 2008. A genotoxicity test system based on p53R2 gene expression in human cells: Assessment of its reactivity to various classes of genotoxic chemicals. *Mutat Res* **656:** 27–35.
Palmer R. 2011. Fecal matters. *Nat Med* **17:** 150–152.
Pharkya P, Burgard AP, Maranas CD. 2004. OptStrain: A computational framework for redesign of microbial production systems. *Genome Res* **14:** 2367–2376.
Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59–65.
Rao S, Hu S, McHugh L, Lueders K, Henry K, Zhao Q, Fekete RA, Kar S, Adhya S, Hamer DH. 2005. Toward a live microbial microbicide for HIV: Commensal bacteria secreting an HIV fusion inhibitor peptide. *Proc Natl Acad Sci* **102:** 11993–11998.
Rasnick D, Duesberg PH. 1999. How aneuploidy affects metabolic control and causes cancer. *Biochem J* **340** (Pt 3)**:** 621–630.
Rinaudo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y. 2007. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol* **25:** 795–801.
Rocha I, Maia P, Evangelista P, Vilaca P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M. 2010. OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Syst Biol* **4:** 45.

Schrattenholz A, Groebe K, Soskic V. 2010. Systems biology approaches and tools for analysis of interactomes and multi-target drugs. *Meth Mol Biol* **662:** 29–58.
Segrè D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci* **99:** 15112–15117.
Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. 2008. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* **26:** 1003–1010.
Smolke CD, Win MN. 2010. Regulating gene expression through engineered RNA technologies. In *The metabolic pathway engineering hand book*. CRC Press, San Diego.
Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, Stahl DA. 2007. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol* **3:** 92.
Strähle U, Klock G, Schütz G. 1987. A DNA sequence of 15 base pairs is sufficient to mediate both glucocorticoid and progesterone induction of gene expression. *Proc Natl Acad Sci* **84:** 7871–7875.
Suhre K, Shin S-Y, Petersen A-K, Mohney RP, Meredith D, Wägele B, Altmaier E, CARDIoGRAM, Deloukas P, Erdmann J, et al. 2011. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477:** 54–60.
Suzuki A, Nakauchi H, Taniguchi H. 2003. Glucagon-like peptide 1 (1–37) converts intestinal epithelial cells into insulin-producing cells. *Proc Natl Acad Sci* **100:** 5034–5039.
Swinburne IA, Miguez DG, Landgraf D, Silver PA. 2008. Intron length increases oscillatory periods of gene expression in animal cells. *Genes Dev* **22:** 2342–2346.
Tang YL, Tang Y, Zhang YC, Agarwal A, Kasahara H, Qian K, Shen L, Phillips MI. 2005. A hypoxia-inducible vigilant vector system for activating therapeutic genes in ischemia. *Gene Ther* **12:** 1163–1170.
Terzer M, Maynard ND, Covert MW, Stelling J. 2009. Genome-scale metabolic networks. *Wiley Interdiscip Rev Syst Biol Med* **1:** 285–297.
Thomas S, Fell DA. 1996. Design of metabolic control for large flux changes. *J Theor Biol* **182:** 285–298.
Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M. 2009. A tunable synthetic mammalian oscillator. *Nature* **457:** 309–312.
Tilg H, Kaser A. 2011. Gut microbiome, obesity, and metabolic dysfunction. *J Clin Invest* **121:** 2126–2132.
Trounson A, Thakar RG, Lomax G, Gibbons D. 2011. Clinical trials for stem cell therapies. *BMC Med* **9:** 52.
Varma A, Palsson BO. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* **60:** 3724–3731.
Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460:** 894–898.
Weber W, Fussenegger M. 2009. Engineering of synthetic mammalian gene networks. *Chem Biol* **16:** 287–297.
Weber W, Link N, Fussenegger M. 2006. A genetic redox sensor for mammalian cells. *Metab Eng* **8:** 273–280.
Win MN, Smolke CD. 2008. Higher-order cellular information processing with synthetic RNA devices. *Science* **322:** 456–460.
Wintermute EH, Silver PA. 2010. Emergent cooperation in microbial metabolism. *Mol Syst Biol* **6:** 407.
Xie Z, Wroblewska L, Prochazka L, Weiss R, Benenson Y. 2011. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333:** 1307–1311.
Ye H, Daoud-El Baba M, Peng R-W, Fussenegger M. 2011. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science* **332:** 1565–1568.
Yearley S. 2009. The ethical landscape: Identifying the right way to think about the ethical and societal aspects of synthetic biology research and products. *J R Soc Interface* **6** (Suppl 4)**:** S559–S564.
Yokoi K, Zhang HS, Kachi S, Balaggan KS, Yu Q, Guschin D, Kunis M, Surosky R, Africa LM, Bainbridge JW, et al. 2007. Gene transfer of an engineered zinc finger protein enhances the anti-angiogenic defense system. *Mol Ther* **15:** 1917–1923.
Yoshida S, Kajimoto Y, Yasuda T, Watada H, Fujitani Y, Kosaka H, Gotow T, Miyatsuka T, Umayahara Y, Yamasaki Y, et al. 2002. PDX-1 induces differentiation of intestinal epithelioid IEC-6 into insulin-producing cells. *Diabetes* **51:** 2505–2513.
Yu RR, Cheng AT, Lagenaur LA, Huang W, Weiss DE, Treece J, Sanders-Beer BE, Hamer DH, Lee PP, Xu Q, et al. 2009. A Chinese rhesus macaque (*Macaca mulatta*) model for vaginal Lactobacillus colonization and live microbicide development. *J Med Primatol* **38:** 125–136.
Zhang Y, Yao L, Shen K, Xu M, Zhou P, Yang W, Liu X, Qin X. 2008. Genetically engineered K cells provide sufficient insulin to correct hyperglycemia in a nude murine model. *Acta Biochim Biophys Sin (Shanghai)* **40:** 149–157.

Appendix E

Personalized Genetic Engineering of Plants[1]

[1] This manuscript has been reviewed for publication and is currently undergoing revisions. The following text is the original unrevised manuscript.

Personalized Genetic Engineering of Plants

Patrick M. Boyle[1*], Devin R. Burrill[1*], Mara C. Inniss[1*], Christina M. Agapakis[1‡*], Aaron Deardon[2†], Jonathan G. DeWerd[2†], Michael A. Gedeon[2†], Jacqueline Y. Quinn[2†], Morgan L. Paull[2†], Anugraha M. Raman[2†], Mark R. Theilmann[2†], Lu Wang[2†], Julia C. Winn[2†], Oliver Medvedik[3], Kurt Schellenberg[4], Karmella A. Haynes[1¶], Alain Viel[3], Tamara J. Brenner[3], Sarah Mathews[4], George M. Church[5,6], Jagesh V. Shah[1§], and Pamela A. Silver[1,5§]

[1] Department of Systems Biology, Harvard Medical School, Boston, MA 02115

[2] Harvard College, Harvard University, Cambridge, MA 02138

[3] Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

[4] The Arnold Arboretum of Harvard University, Boston, MA 02131

[5] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115

[6] Department of Genetics, Harvard Medical School, Boston, MA 02115

[*] These authors contributed equally to this work

[†] These authors contributed equally to this work

[§] Corresponding authors

[‡] Current Address: Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095

[¶] Current Address: School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287

Email addresses:

PMB : pmboyle@fas.harvard.edu

DRB: dburrill@fas.harvard.edu

MCI: minniss@fas.harvard.edu

CMA: agapakis@ucla.edu

AD: deardon@post.harvard.edu

JGdW: jjoonathan@gmail.com

MAG: gedeon@fas.harvard.edu

JYQ: jyquinn@gmail.com

MLP: mpaull15@gmail.com

AMR: amraman1991@gmail.com

MRT: mtheilm@fas.harvard.edu

LW: lu.irena.wang@gmail.com

JCW: unfinished.art@gmail.com

OM: omedvedik@gmail.com

KS: kurtschelle@gmail.com

KAH: karmella.haynes@asu.edu

AV: aviel@fas.harvard.edu

TJB: tamara_brenner@harvard.edu

SM: smathews@oeb.harvard.edu

GMC: gmc@harvard.edu

JVS[§]: jagesh_shah@hms.harvard.edu

PAS[§]: pamela_silver@hms.harvard.edu

Abstract

Background

Plant biotechnology can be leveraged to produce food, fuel, medicine, and materials. Standardized methods advocated by the synthetic biology community can accelerate the plant design cycle; ultimately making plant engineering more widely accessible to bioengineers who can contribute diverse creative input to the design process.

Results

This paper presents work done largely by undergraduate students participating in the 2010 International Genetically Engineered Machines (iGEM) competition. Described here is a framework for engineering the model plant *Arabidopsis thaliana* with open-source, standardized vectors and parts available through the MIT Registry of Standard Biological Parts (www.partsregistry.org). This system was used to engineer a proof-of-concept plant that expresses a taste-inverting exogenous protein, miraculin.

Conclusions

Collectively, our work is intended to facilitate plant synthetic biology. We envision a future where plant biotechnology can be designed by and tailored to the needs of the consumer at local rather than industrial scales. This kind of *personalized* genetic engineering can potentially decrease the environmental cost of agriculture, enhance foods' nutritional value, and increase public understanding of genetically modified organisms.

Keywords

iGEM, synthetic biology, Arabidopsis, plant biotechnology, miraculin

Background

Humans have long used selective breeding to modify plant characteristics such as growth rate, seed size, and flavor [1]. For much of agricultural history, the targeted traits reflected the needs of local growers and consumers, creating a vast array of crop varieties. Advances in the field of genetics and the advent of recombinant DNA technology accelerated our ability to manipulate food crops [1-5]. The introduction of multiple genes (termed *gene stacking* in plants) has made plants accessible to synthetic biology applications [6-11].

Despite these advances, industrial interests and gene patents have significantly limited the development of diverse platforms for plant engineering methods and applications. Monocultural approaches are widely favored in industrial agriculture and this monopolizes agricultural engineering, which threatens food security, public health, and ecosystem stability. Of the thousands of edible plant species, just five contribute to the majority of agricultural output: wheat, rice, soy, corn, and rape [12]. Furthermore, the number of industrial patents claimed on protocols related to *Agrobacterium tumefaciens*, the gram-negative soil bacterium widely used for plant genetic engineering, has significantly increased over the last thirty years [13]. The labyrinth of patent and intellectual property rights for plant biotechnological processes and materials, including *Agrobacterium*-mediated protocols, creates an obstacle to the private and public use of plant engineering tools.

This work aims to create an open-source, standardized, and modular system for the production of genetically enhanced plants to facilitate their adoption by diverse users outside of industry and academia. Ideally, an open-source plant engineering system is customizable, yet has convenient standard features that minimize the need to re-invent common steps such as transferring genetic material into the plant. We demonstrate the feasibility of small-scale, personalized engineering projects in the model organism, *Arabidopsis thaliana* (Arabidopsis), using a BioBrick-modified plant vector system (Figure 1). As a proof-of-principle, Arabidopsis was modified to express miraculin, a flavor-inverting protein from *Synsepalum*

dulcificum that is consumed as a novelty product. Beyond creative gastronomy, we imagine this system being used to enhance the nutritional content of edible plants or help allergy sufferers enjoy the benefits of fresh home-grown produce (Figure 1). The development of efficient transformation techniques for many plants remains a key hurdle for commercial and personal agriculture. However, flexible genetic customization of plants also requires a system of easily transferable, standardized components such as those presented here. We hope this work will lead to techniques that yield a diversity of produce tailored to individual, community, and local environmental needs.

Results

Design of BioBrick compatible vectors for Arabidopsis transformation

Arabidopsis is readily transformed by *Agrobacterium*: when a plant is injured, *Agrobacterium* localizes to the wound site and naturally transfers the T-DNA region of its tumor-inducing (Ti) plasmid into the plant cell [14]. The T-DNA localizes to the nucleus and integrates into the plant's chromosomal DNA. A series of open-source vectors (the pORE series) has been developed from *Agrobacterium*'s Ti plasmid to allow transformation of heterologous DNA into plants via *Agrobacterium* [14]. pORE vectors come equipped with a multiple cloning site (MCS) containing twenty-one unique restriction endonuclease sites. Reporters or promoters are included to create expression vectors, reporter vectors, or vectors that can carry an exogenous promoter or open reading frame. This vector series offers either glufosinate resistance via the *pat* gene, or kanamycin resistance via the *nptII* gene, to enable the selection of successfully transformed plants.

We developed a new set of six BioBrick DNA assembly-compatible plant transformation vectors based on the pORE series (Table 1). Vectors V1 and V2 (modified Open vectors) contain no promoter or reporter gene, allowing integration of constructs under

the control of a chosen promoter (Figure 2A, Table 1). Vectors V3 and V4 (modified Expression vectors) contain the constitutive pENTCUP2 promoter upstream of the MCS (Figure 2B, Table 1), while V5 and V6 (modified Reporter vectors) contain no promoter but have either the reporter gusA or soluble modified GFP (smgfp) downstream of the cloning site (Figure 2C, Table 1). Each vector contains an MCS that is compatible with three widely used BioBrick standards (RFC 10, 20, 23, www.partsregistry.org).

Expression of standardized flavor protein genes in industrial microorganisms

Using the Biobricked vectors, we sought to modify the taste of Arabidopsis, specifically enhancing the sweetness of a bitter plant without altering sugar content. Several naturally occurring proteins are 100 – 3000 times sweeter than sugar by weight [15]. Brazzein, monellin, thaumatin, pentadin, mabinlin, and curculin are sweet proteins found in a variety of African and South Asian fruits, with no sequence similarity or common features [16]. Brazzein, isolated from the West African fruit *Pentadiplandra brazzeana*, is the smallest of these proteins with only 54 amino acids. It has a high heat tolerance, surviving 80° C exposure for several hours, and was previously expressed heterologously in *Escherichia coli* (*E. coli*) [16], *Zea mays* [17], and *Lactobacillus lactis* [18].

Miraculin, isolated from the berries of the West African plant *Richadella dulcifica*, does not taste sweet on its own. Rather, it acts as a flavor-inverter by binding to taste receptors on the tongue in a pH-dependent manner, causing sour foods to taste sweet [19]. A 1 μM miraculin solution is sufficient to activate this inversion, where 20 mM citrate corresponds to the sweetness of 300 mM sucrose [20]. Miraculin is a glycosylated homodimer that has been heterologously expressed in lettuce [21], tomato [22], and even *E. coli* [20], indicating that endogenous *Richadella* glycosylation is not required for functional expression.

Full-length miraculin and brazzein protein expression from standardized open reading frames was tested in *E. coli* and the yeast *Saccharomyces cerevisiae*. Brazzein and miraculin genes were commercially synthesized and codon-optimized for expression in Arabidopsis. Biobrick-compatible restriction enzyme sites bracketed each open reading frame. Constructs were tagged at either the N- or C-terminus with the Strep-II tag [23] for western blot analysis. Miraculin (Figure 3A) and brazzein (Figure 3B) were expressed from an IPTG inducible T7 promoter in *E. coli*. Monomeric miraculin was expressed at very low levels at approximately 24 kDa regardless of tag location, which is consistent with previous work [20]. Brazzein was highly expressed in the same system at about 12 kDa, regardless of the tag location, which is also consistent with previous results [16]. Brazzein was also highly expressed from the constitutive TEF and copper-inducible CUP1 promoters in yeast (Figure 3C). The higher molecular weight of the Strep-II tagged brazzein observed by western blot in yeast, compared to *E. coli* (~35 kDa versus 12 kDa) is likely due to yeast-specific glycosylation of the brazzein protein[24]. While expression of the miraculin gene was not verified in yeast we attempted integration of both miraculin and brazzein constructs in Arabidopsis.

Expression of flavor proteins in Arabidopsis

We successfully introduced two different BioBrick plant vectors into Arabidopsis and selected for seeds carrying genomically-integrated miraculin and brazzein transgenes. Miraculin- or brazzein-encoding DNA was introduced into Arabidopsis via *Agrobacterium*-mediated transformation [25] under control of the pENTCUP2 promoter and the NosT transcriptional terminator on either the V3 (glufosinate resistance) or V4 (kanamycin resistance) BioBrick vector. Transformed seeds were selected on MS-agar, and resistant plants were moved to soil and allowed to produce seeds. T1 generation seeds were collected and re-plated on selective plates. Resistant plants were once again moved to soil and allowed to produce T2 generation seeds. While integration of both the miraculin and brazzein genes

into the plant genome was verified by PCR (Figure 4A), only miraculin RNA expression was detected (Figure 4B). Furthermore, miraculin protein expression was verified by western blot on whole cell protein (Figure 5). A band was detected at ~ 40 - 45 kDa that was not present in a protein sample from a control plant of the same background strain; this protein size is consistent with that previously observed for a glycosylated miraculin monomer in lettuce [21]. The antibody shows significant background binding (see Additional file 1, Figure S1), however, taken together (with the RNA expression data) these data indicate that the miraculin protein is expressed in our transgenic plants. We present a unique application of open-source, standardized, modular parts to engineer plants, establishing a new avenue for the personalized genetic enhancement of non-commercial agriculture.

Discussion and Conclusions

Genetic engineering of plants at the industrial scale is well established. Technological advances have yielded crops that reduce food production costs through resistance to pests, herbicide, drought, and flood [26]. Additionally, modification of crops (e.g., rice) to contain pro-vitamins can help treat health issues such as vitamin A deficiency in countries where staple foods do not provide the necessary nutrients [26, 27]. However, small-scale experimental horticulture, farming, and gardening have been excluded from producing innovative contributions due to the lack of accessible tools for the genetic modification of plants. The availability of standardized plant vectors in the Registry of Standard Biological Parts will facilitate the development of small-scale plant engineering projects.

We have modified existing open-source plant integration vectors to make them compatible with the BioBrick assembly standard 23 [28], demonstrated that they can be used to integrate transgenes in Arabidopsis, and showed successful integration and expression of the taste modifying protein miraculin. All constructs have been submitted to the Registry of

Biological Parts (www.partsregistry.org) and are available as a resource for the synthetic biology and plant engineering community. These include vectors modified from the pORE vector series [14], plant specific regulatory elements (e.g., promoters, terminators), resistance markers, and the coding sequence of miraculin. The vector series features variations containing the constitutive promoter pENTCUP2 (V3 and V4), visible reporters gusA or smGFP (V5 and V6), or a simple multiple cloning site (V1 and V2), allowing expression of a gene from a promoter of choice.

In addition to using these vectors to express exogenous proteins, we have considered integrating constructs expressing hairpin RNAs [29] or artificial microRNAs [30] to knock down the expression of endogenous genes. This strategy is particularly powerful in that synthesizing a DNA sequence to match any gene transcript of choice allows the regulation of potentially any plant protein. For instance, this approach could be used reduce allergenic protein levels [31]. Alternatively, microRNAs could be targeted to metabolic regulators so that key metabolites, such as pigments or nutrients, are allowed to accumulate [32] and enhance the color or nutritional content of the plant. Modification of existing open-source vectors to conform to the BioBrick assembly standard allows them to be integrated into a BioBrick cloning based workflow. In addition to simplifying the construction of more complex genetic devices, adhering to an assembly standard allows for the possibility of automation of assembly.

We hope that availability of plant integration vectors compatible with a common assembly standard will facilitate the use of plants as a chassis in synthetic biology. This work demonstrates the feasibility of customizing the genetic makeup of plants with open-source and standardized components. Garden-scale design of food plants, in which the gardener selects desirable traits, can be made possible through the availability of low-cost or free genetic parts. Personalized engineering of plants to modify flavor, nutritional value, or

allergenicity could create a new class of synthetic organic foods that are grown and consumed at a local scale.

Materials and methods

Plasmids and cloning

Gene assembly was performed in *E. coli* DH5α using BioBrick assembly standard 21 [28], and all described parts were submitted to the BioBrick Registry. Arabidopsis pORE series vectors were provided by The Arabidopsis Information Resource (TAIR) and engineered to support BioBrick cloning through PCR-based methods (see Additional file 1, Table S1). pORE Open Series vectors O1 and O2 were digested with SpeI and SacII and ligated with an annealed oligonucleotide insert with NheI and SacII overhangs containing the BioBrick Multiple Cloning Site (MCS) to create vectors V1 and V2. pORE Expression Series vectors E3 and E4 were likewise digested with HindIII and SpeI and ligated with an insert PCR amplified from the expression vectors containing HindIII site upstream from the pENTCUP2 promoter and the BioBrick MCS and an NheI site downstream to create vectors V3 and V4. pORE Reporter Series vectors R1, containing the gusA reporter, and R3, containing the smgfp reporter, were digested with HindIII and SpeI and ligated with inserts containing the reporter gene PCR amplified with primers containing a HindIII site followed by the BioBrick MCS upstream and NheI downstream, yielding vectors V5 and V6.

Brazzein and miraculin were codon optimized for expression in Arabidopsis, commercially synthesized (Mr. Gene, Regensburg, Germany), and assembled with the pENTCUP2 promoter and NosT transcriptional terminator. Completed constructs were subcloned from BioBrick assembly vector V0120 to BioBrick modified pORE vectors through digestion with EcoRI and PstI.

Plant Maintenance

Wild-type Col-0 *Arabidopsis thaliana* seeds were sterilized by washing with 70% ethanol, 0.1% Triton X-100, followed by two 95% ethanol washes and two sterile dH_2O washes. Seeds were then plated on 1X Murashige & Skoog (MS) media with 0.7% agar supplemented with 150 uM carbenicillin and placed in the dark at 4 ºC for three days before moving to an incubator with 16 h illumination at 20 ºC and 8 h dark at 15 ºC per day to allow seeds to germinate. Once plants produced secondary leaves, they were moved to soil and allowed to mature and produce seeds. Seeds were collected and stored at 4 ºC.

Plant Transformation

Agrobacterium-mediated transformation was performed according to previously reported techniques [25]. Briefly, *Agrobacterium* was made electro-competent by washing in cold sterile water and resuspending in 10% glycerol. Vector DNA was dialyzed to remove excess salt, and electroporated into *Agrobacterium*. Kanamycin resistant colonies were grown in YEB media, spread on YEB plates, and allowed to form a lawn. Lawns were scraped and suspended in a solution of 20% YEB, 4% sucrose (w/v), and 0.024% Silwet L-77 surfactant (Helena Chemical Company, Collierville, TN). Wild-type Col-0 Arabidopsis flowers were dipped in the *Agrobacterium* solution and allowed to grow and develop seed pods. Seeds were collected from mature plants and selected on 1x MS media with 0.7% agar supplemented with 5 mg/L glufosinate or 50 μg/ml kanamycin.

***E. coli* and yeast protein expression**

In BL21(DE3) *E. coli*, StrepII-tagged brazzein and miraculin were inserted at multiple cloning site 1 of a BioBrick-modified pET-duet vector [33]. Cells were grown to mid-log phase and induced with a final concentration of 1 mM IPTG. Protein expression was measured by western blot.

In PSY580a yeast *S. cerevisiae*, StrepII-tagged brazzein was cloned with the constitutive TEF promoter or the copper-inducible CUP1 promoter and integrated at the LEU2 locus. Transformants were grown in YEPD media with 0.3 mM $CuSO_4$ to induce protein expression, which was measured by western blot.

Verification of genomic transgenes

Genomic DNA was extracted using the DNEasy kit (Qiagen) and amplified by PCR (see Additional file 1, Table S2). Whole cell DNA Whole cell RNA was collected using the plant RNEasy kit (Qiagen). cDNA was synthesized with the SuperScript III First-Strand synthesis kit (Invitrogen). qPCR was performed with primer pairs (see Additional file 1, Table S2) amplifying 100 base pair amplicons within target genes to identify expression of heterologous genes or endogenous gene knockdown.

SDS-Page and Western Blotting

Protein samples were extracted from Arabidopsis, *E. coli*, and yeast and normalized using the Bradford assay (Bio-Rad, Hercules, CA). Samples were diluted into SDS-PAGE loading buffer and loaded onto a 4-20% Tris/glycine/SDS acrylamide gel. α-Strep-tag II antibody (HRP-conjugated, Novagen, Gibbstown, NJ) was used to measure brazzein and miraculin protein expression in yeast and *E. coli*, and α-miraculin antibody [21] (provided by Tadayoshi Hirai, Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan) was used to detect levels of miraculin expression in Arabidopsis. Monoclonal Anti-β-Tubulin antibody (Sigma-Aldrich, St-Louis, MO) was used to detect tubulin in Arabidopsis.

List of abbreviations used

MCS: multiple cloning site

Competing interests

The authors declare they have no competing interests.

Authors' contributions

Cloning schemes were designed by PMB, DRB, MCI, CMA, AD, JGdW, AAG, JYQ, MLP, AMR, MRT, LW, JCW, and OM. KS and SM provided technical assistance with Arabidopsis culture and *Agrobacterium*-mediated transformation. PMB, DRB, MCI, and CMA performed PCR, qPCR and western blots; all other cloning and experiments were performed by AD, JGdW, AG, JYQ, MLP, AMR, MRT, LW, and JCW. KAH, AV, TJB, SM, GMC, JVS, and PAS provided general advising throughout the project. The manuscript was drafted by PMB, DRB, MCI, and CMA. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the people at iGEM for supporting this project and allowing us the opportunity to present it at the 2010 iGEM competition. PMB was supported by the Harvard University Center for the Environment Graduate Consortium. PMB and DRB were supported by NSF SYNBERC (520.45321.148830.377227.0402.66376). MCI was supported by the Natural Science and Engineering Research Council of Canada (NSERC). CMA was supported by a NSF Graduate Research Fellowship. AD, JGdW, MAG, JYQ, MLP, AMR, MRT, LW, JCW, and OM were generously supported by the Wyss Institute for Biologically Inspired Engineering and a grant from the Howard Hughes Medical Institute Undergraduate Education Program awarded to Robert A. Lue, Department of Molecular and Cellular Biology, Harvard University. KS was supported by the Arnold Arboretum.

References

1. Kingsbury N: *Hybrid: the history and science of plant breeding - Noël Kingsbury - Google Books*. University of Chicago Press; 2009.

2. Weiling F: *Historical study: Johann Gregor Mendel 1822-1884.* 1991, **40**:1–25.

3. Jackson DA, Symons RH, Berg P: **Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli.** *Proc Natl Acad Sci USA* 1972, **69**:2904–2909.

4. Lobban PE, Kaiser AD: **Enzymatic end-to end joining of DNA molecules.** *J Mol Biol* 1973, **78**:453–471.

5. Cohen SN, Chang AC, Boyer HW, Helling RB: **Construction of biologically functional bacterial plasmids in vitro.** *Proc Natl Acad Sci USA* 1973, **70**:3240–3244.

6. Antunes MS, Ha S-B, Tewari-Singh N, Morey KJ, Trofka AM, Kugrens P, Deyholos M, Medford JI: **A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity**. *Plant Biotechnology Journal* 2006, **4**:605–622.

7. Herrera-Estrella L, Depicker A, Van Montagu M, Schell J: **Expression of chimaeric genes transferred into plant cells using a Ti-plasmid-derived vector**. *Nature* 1983, **303**:209–213.

8. Bevan MW, Flavell RB, Chilton M-D: **A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation**. *Nature* 1983, **304**:184–187.

9. Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL, Woo SC: **Expression of bacterial genes in plant cells.** *Proc Natl Acad Sci USA* 1983, **80**:4803–4807.

10. Murai N, Kemp JD, Sutton DW, Murray MG, Slightom JL, Merlo DJ, Reichert NA, Sengupta-Gopalan C, Stock CA, Barker RF, Hall TC: **Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors.** *Science* 1983, **222**:476–482.

11. Shewry PR, Jones HD, Halford NG: *Advances in Biochemical Engineering/Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008, **111**:149–186.

12. Kern M: **Food, Feed, Fibre, Fuel and Industrial Products of the Future: Challenges and Opportunities. Understanding the Strategic Potential of Plant Genetic Engineering**. *J Agron Crop Sci* 2002, **188**:291–305.

13. **Agrobacterium-mediated transformation of plants** [http://www.bios.net/daisy/bios/81/version/live/part/4/data].

14. Coutu C, Brandle J, Brown D, Brown K, Miki B, Simmonds J, Hegedus DD: **pORE: a modular binary vector series suited for both monocot and dicot plant transformation**.

Transgenic Res 2007, **16**:771–781.

15. Kant R: **Sweet proteins--potential replacement for artificial low calorie sweeteners.** *Nutr J* 2005, **4**:5.

16. Assadi-Porter F: **Efficient Production of Recombinant Brazzein, a Small, Heat-Stable, Sweet-Tasting Protein of Plant Origin**. *Arch Biochem Biophys* 2000, **376**:252–258.

17. Lamphear BJ, Barker DK, Brooks CA, Delaney DE, Lane JR, Beifuss K, Love R, Thompson K, Mayor J, Clough R, Harkey R, Poage M, Drees C, Horn ME, Streatfield SJ, Nikolov Z, Woodard SL, Hood EE, Jilka JM, Howard JA: **Expression of the sweet protein brazzein in maize for production of a new commercial sweetener**. *Plant Biotechnology Journal* 2004, **3**:103–114.

18. Berlec A, Jevnikar Z, Majhenič AČ, Rogelj I, Štrukelj B: **Expression of the sweet-tasting plant protein brazzein in Escherichia coli and Lactococcus lactis: a path toward sweet lactic acid bacteria**. *Appl Microbiol Biotechnol* 2006, **73**:158–165.

19. Koizumi A, Tsuchiya A, Nakajima K-I, Ito K, Terada T, Shimizu-Ibuka A, Briand L, Asakura T, Misaka T, Abe K: **Human sweet taste receptor mediates acid-induced sweetness of miraculin.** *Proc Natl Acad Sci USA* 2011, **108**:16819–16824.

20. Matsuyama T, Satoh M, Nakata R, Aoyama T, Inoue H: **Functional Expression of Miraculin, a Taste-Modifying Protein in Escherichia Coli**. *Journal of Biochemistry* 2009, **145**:445–450.

21. Sun H-J, Cui M-L, Ma B, Ezura H: **Functional expression of the taste-modifying protein, miraculin, in transgenic lettuce.** *FEBS Letters* 2006, **580**:620–626.

22. Hirai T, Fukukawa G, Kakuta H, Fukuda N, Ezura H: **Production of Recombinant Miraculin Using Transgenic Tomatoes in a Closed Cultivation System**. *J. Agric. Food Chem.* 2010, **58**:6096–6101.

23. Schmidt TG, Koepke J, Frank R, Skerra A: **Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin.** *J Mol Biol* 1996, **255**:753–766.

24. Carlson A, Armentrout RW, Ellis TP: **Enhanced Production and Purification of a Natural High Intensity Sweetener**. 2010:1–34.

25. Logemann E, Birkenbihl RP, Ulker B, Somssich IE: **An improved method for preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol**. *Plant Methods* 2006, **2**:16.

26. Ronald P: **Plant Genetics, Sustainable Agriculture and Global Food Security**. *Genetics* 2011, **188**:11–20.

27. Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I: **Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm.** *Science* 2000, **287**:303–305.

28. Phillips IE, Silver PA: **A New Biobrick Assembly Strategy Designed for Facile Protein Engineering**. 2006:1–6.

29. Helliwell C, Waterhouse P: **Constructs and methods for high-throughput gene silencing in plants.** *Methods* 2003, **30**:289–295.

30. Ossowski S, Schwab R, Weigel D: **Gene silencing in plants using artificial microRNAs and other small RNAs.** *Plant J* 2008, **53**:674–690.

31. Singh MB, Bhalla PL: **Genetic engineering for removing food allergens from plants.** *Trends Plant Sci* 2008, **13**:257–260.

32. Kennedy CJ, Boyle PM, Waks Z, Silver PA: **Systems-level engineering of nonfermentative metabolism in yeast.** *Genetics* 2009, **183**:385–397.

33. Agapakis CM, Ducat DC, Boyle PM, Wintermute EH, Way JC, Silver PA: **Insulation of a synthetic hydrogen metabolism circuit in bacteria.** *J Biol Eng* 2010, **4**:3.

Figure Legends

Figure 1 - An open-source, standardized, and modular system for the production of genetically-modified plants.

Genetic parts obtained from the BioBrick Registry were assembled and inserted into modified vectors (Open, Expression, or Reporter) in *E. coli*. These parts may be assembled to build constructs to impact a wide variety of plant phenotypes. Once assembled, these vectors were transformed into *Agrobacterium*. Using the floral dip procedure, Agrobacterium infected Arabidopsis, thereby transferring the assembled construct. Once seeds were produced, they were plated on selective media to obtain transgenic plants carrying the assembled construct.

Figure 2 - Schematic of BioBrick plant vectors.

(A) Modified Open vectors are based on TAIR vectors pORE O1 and O2 [14]. They are designed for general insertion of a construct. (B) Modified Expression vectors are based on TAIR vectors pORE E3 and E4 [14]. They contain an inducible promoter preceding the BioBrick MCS, to permit user-controlled expression of the inserted construct. (C) Modified Reporter vectors are based on TAIR vectors pORE R1 and R2 [14]. They contain a reporter gene following the BioBrick MCS, such that expression of the reporter follows that of the inserted construct.

Figure 3 - BioBrick miraculin and brazzein protein expression in bacteria and yeast.
(A) Miraculin and (B) brazzein BioBricks were expressed from an IPTG-inducible promoter in *E. coli* with an N- or C-terminal Strep-II tag . (C) Brazzein BioBrick was expressed in yeast from pTEF or pCup1 promoters C- terminal Strep-II tag.

Figure 4 - BioBrick miraculin DNA and RNA expression in Arabidopsis.
(A) Integration of the miraculin and brazzein genes in the Arabidopsis genome was confirmed. (B) Miraculin RNA was constitutively expressed in Arabidopsis however brazzein expression was not detected. act: actin control; b: brazzein primer set; m1-m3; miraculin primer sets.

Figure 5 - BioBrick miraculin protein is expressed in Arabidopsis.
Miraculin was constitutively expressed in transgenic plant, as compared to control plant (WT). Closed arrow: miraculin band. Open arrows: non-specific bands.

Tables

Table 1 – Features of BioBrick plant vectors

Vector	Biobrick Registry ID	Bacterial Resistance	Plant Resistance	Promoter	Reporter	Original pORE vector
V1	BBa_K382000	Kan	Pat	none	none	pORE O1
V2	BBa_K382001	Kan	nptII	none	none	pORE O2
V3	BBa_K382002	Kan	Pat	pENTCUP2	none	pORE E3
V4	BBa_K382003	Kan	nptII	pENTCUP2	none	pORE E4
V5	BBa_K382004	Kan	nptII	none	gusA	pORE R1
V6	BBa_K382005	Kan	nptII	none	smgfp	pORE R3

Additional files

Additional file 1 – Supplementary Data for Personalized Genetic Engineering of Plants.

Contains tables of primer sequences describing primers used to modify pORE vectors, verify integration of transgenes, and RNA expression as well as the uncropped western blot shown in Figure 5.

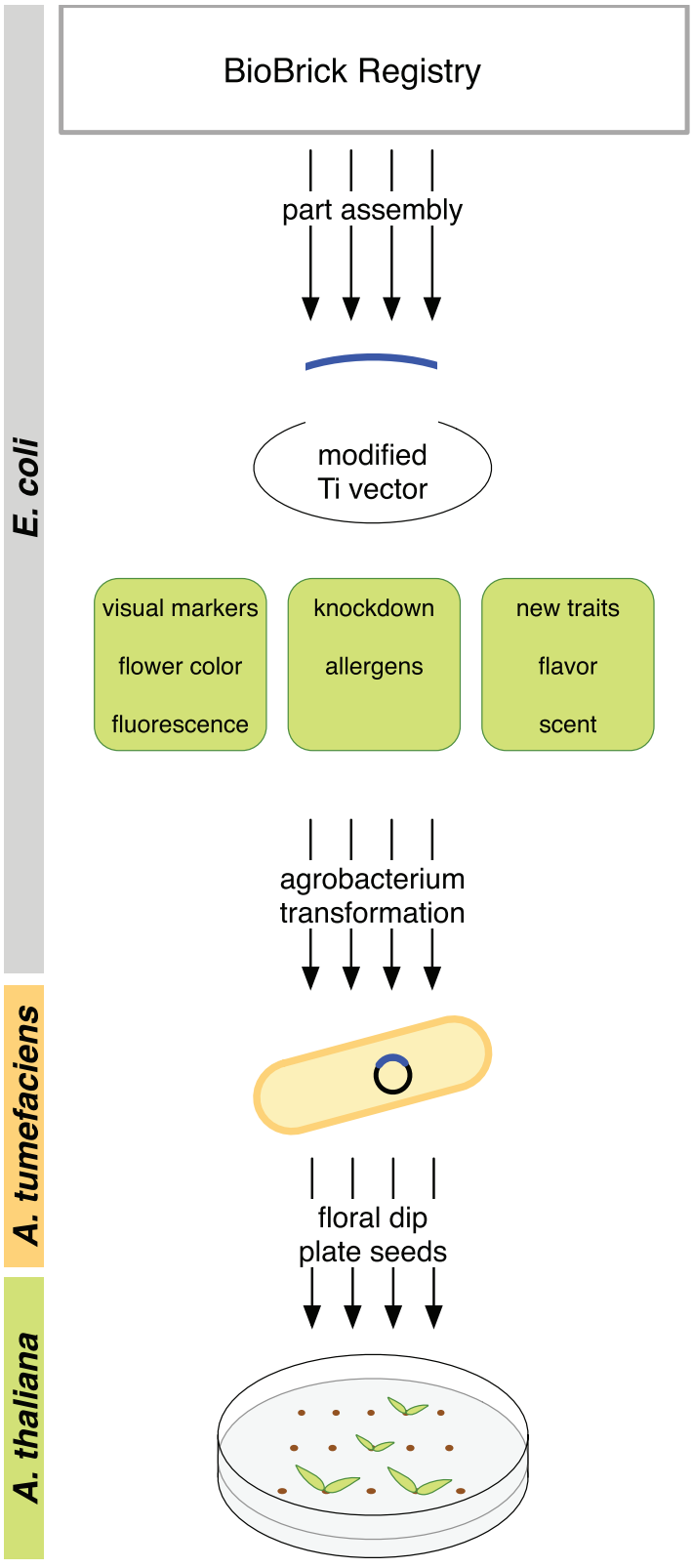

E. coli
A. tumefaciens
A. thaliana
BioBrick Registry
part assembly
modified
Ti vector
visual markers
flower color
fluorescence
knockdown
allergens
new traits
flavor
scent
agrobacterium
transformation
floral dip
plate seeds

Figure 1

Boyle et al. Figure 2

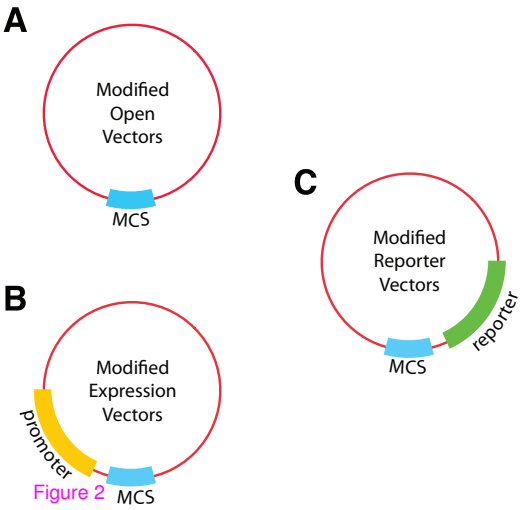


Figure 2

Boyle et al. Figure 3

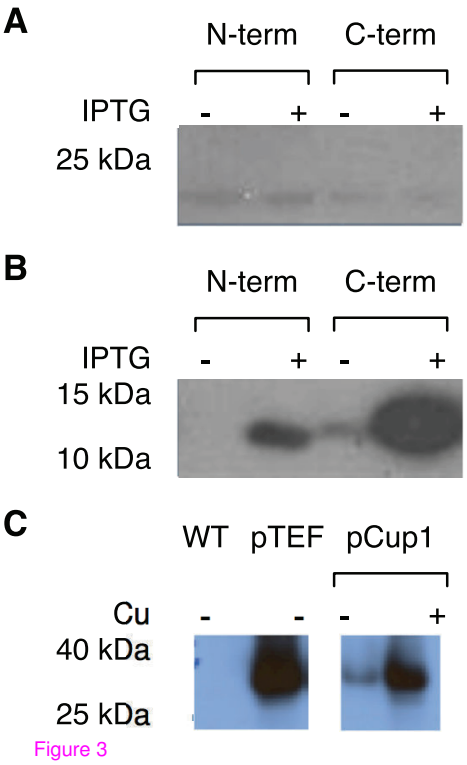


Figure 3

Boyle et al. Figure 4

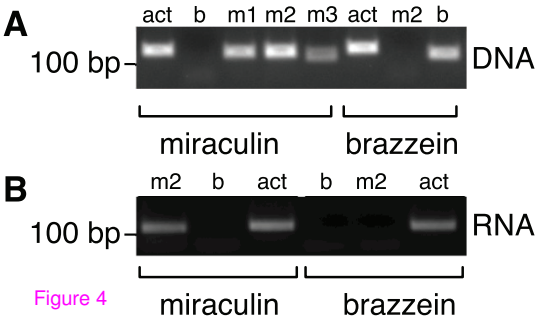


Figure 4

Boyle et al. Figure 5

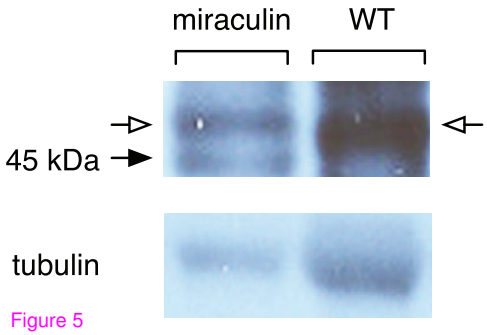


Figure 5

Appendix F

Natural strategies for the spatial optimization of metabolism in synthetic biology[1]

[1] This manuscript was written as an invited review for Nature Chemical Biology, and has been accepted for publication. The following is the final preprint draft of the manuscript, and is reproduced with permission from the Editor.

Natural strategies for the spatial optimization of metabolism in synthetic biology

Christina M. Agapakis[1*], Patrick M. Boyle[2*], and Pamela A. Silver[2,3§]

[1] Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095

[2] Department of Systems Biology, Harvard Medical School, Boston, MA 02115

[3] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115

* These authors contributed equally to this work

[† §] Corresponding author: pamela_silver@hms.harvard.edu

Abstract

Metabolism is a highly interconnected web of chemical reactions that power life. While the stoichiometry of metabolism is well understood, the multi-dimensional aspects of metabolic regulation in time and space remain difficult to define, model, and engineer. Complex metabolic conversions can be performed by multiple species working cooperatively, exchanging metabolites via structured networks of organisms and resources. Within cells, metabolism is spatially regulated as well, via sequestration in subcellular compartments and through the assembly of multi-enzyme complexes. Metabolic engineering and synthetic biology have had success in engineering metabolism in the first and second dimensions, designing linear metabolic pathways and channeling metabolic flux. More recently, engineering of the third dimension has improved output of engineered pathways through isolation and organization of multi-cell and multi-enzyme complexes. This review highlights natural and synthetic examples of three dimensional metabolism both inter- and intra-cellularly, offering tools and perspectives for biological design.

Cellular metabolism can accomplish chemical feats at specificities, temperatures, and pressures that chemists and chemical engineers can only dream of. Synthetic biology and metabolic engineering seek to turn living cells into microbial factories: self-regenerating machines producing renewable fuels, medicines, and materials. Learning from natural strategies for solving complex metabolic problems can significantly optimize the behavior of these biotechnologies. In this review we discuss how natural and synthetic metabolic pathways can be optimized through spatial organization, from the cell scale across microbial consortia, to the protein scale inside cellular enzyme complexes. Our story begins inside the cockroach hindgut, where a remarkable methane production system shows the importance of multi-scale spatial regulation in enzymatic pathways.

The cockroach gut may seem like a strange place to start a review about cellular metabolism, but cockroaches contain a multilayered and surprisingly structured metabolism that allows them to digest and survive off of an extraordinarily wide range of food sources, including cellulose-rich materials such as paper[1]. Cellulose degradation requires the coordinated behavior of many different enzymes and multiple cellular pathways, and indeed only a relatively small number of microbial organisms have evolved the necessary enzymatic systems. Herbivorous animals, including cockroaches, thus digest cellulose by outsourcing the breakdown chemistry to a complex community of these microbes—protozoa, archaea, bacteria, and fungi in the bovine rumen or insect hindgut that perform a series of linked chemical reactions (Fig. 1). Microbial communities living in the hindgut of cockroach species such as *Periplaneta americana*

can anaerobically digest cellulose, using carbon dioxide as a terminal electron acceptor to generate up to 4ml of methane per insect per day[1].

Investigating this methane production revealed an unexpected source: the anaerobic eukaryote *Nyctotherus ovalis. N. ovalis*, is required for healthy cockroach growth and present at high cell densities (up to 5-6 x 10^4 cells/ml) in the hindgut[1]. Methanogenesis is rare in biology and had only been observed in archaea. Further unraveling of this system explains that gut-dwelling ciliates such as *N. ovalis* do not produce methane on their own, but instead harbor methanogenic endosymbionts—archaea of the genus *Methanobrevibacter* that complete the final steps of the chemical breakdown of cellulose[1-3]. These vertically-transmitted endosymbionts are closely related to free-living species, indicating multiple acquisition events[2].

This nested endosymbiosis maintains the thermodynamic equilibrium of the whole community and significantly impacts the fermentative ability of *N. ovalis*[4]. Methanogens are autotrophs that consume hydrogen as a source of reducing power to fix carbon dioxide, while *N. ovalis* couples ATP generation to hydrogen production in the hydrogenosome, an organelle evolutionarily related to the mitochondria[5]. By decreasing the partial pressure of hydrogen outside the hydrogenosomes, the methanogens ensure that the hydrogen-producing reactions continue to be thermodynamically favorable. Spatial proximity is crucial to the ability of the methanogens to efficiently siphon hydrogen from the hydrogenosome, with electron micrographs showing the methanogenic endosymbionts tightly associated with the hydrogenosomes' outer surface[3] (Fig. 1).

Spatial optimization: learning from nature

The spatial coupling of complex metabolic reactions is a common theme throughout biology, with reactions isolated from one another in different cells, compartments, or complexes and subsequently linked through controlled proximity. Three-dimensional organization and optimization at many scales can concentrate reactants to drive unfavorable reactions[6], remove inhibitory products[7], or channel metabolites from one enzyme to the next[8]. These strategies have been exploited in metabolic engineering to improve pathway function[9], and will likely play an important role in future attempts to design metabolic pathways for the breakdown and/or production of complex organic molecules.

The optimization of synthetic metabolic pathways has been described as a zero-sum game, where channeling flux to an artificial pathway necessarily takes energy away from required cellular networks[10]. Microbial consortia could attenuate these challenges through division of labor among a diverse population, driving thermodynamically unfavorable reactions by creating concentration gradients between species, as well as physically isolating toxic enzymes or intermediates into separate cell types[11]. Furthermore, consortia may be able to more efficiently perform consolidated bioprocesses than a single organism, in particular the breakdown of cellulose and subsequent production of biofuels[12-14]. Natural strategies and mechanisms that govern the association of microbes in symbiotic assemblages can influence the way that synthetic metabolic pathways are designed across multiple species.

Inside cells, metabolic networks function amidst chemical chaos using a number of strategies that bring together appropriate cells, enzymes, and substrates in time and space. These optimization strategies function at many scales and are responsive to external signals, ensuring that the cell does not waste metabolic resources by producing unnecessary enzymes, as well as channeling flux through proper metabolic routes. Cellular mechanisms that scaffold or otherwise spatially isolate metabolic pathways can ensure proper metabolic function in a number of interconnected ways. First, metabolites or enzymes that can react promiscuously can be channeled and scaffolded to maintain specificity[15]. Second, channeling[16], scaffolding[8], or compartmentalization[17] can concentrate reactants to drive unfavorable reactions, protect enzymes or unstable intermediates from harmful cellular conditions or competing reactions, as well as protect the rest of the cell from toxic intermediates.

Many of these regulatory strategies are modular, whether they are multi-domain or multi-polypeptide enzyme complexes[18], scaffolded or linked with interchangeable protein-protein interaction domains, or whether they function at the level of whole organelles or cells. The modularity of natural spatial organization of metabolism enables these pathways to be robust as well as evolutionarily flexible[19], new reactions and pathways can rapidly evolve through the establishment of new symbiotic relationships or the recombination of protein domains.

In this review, we explore evolved and engineered metabolic optimization strategies from the multicellular to the protein scale. We emphasize the modular biological components that comprise these strategies, as synthetic biologists seek to

repurpose these components for the improvement of novel pathways. Multicellular complexes can be modulated with the addition of new species, as well as designed from scratch and spatially controlled through microfluidics. Inside cells, targeting of proteins to compartments such as the carboxysome, vacuole, or peroxisome can concentrate reactions and pathways that need special cellular conditions to function properly and may be toxic in the cytoplasmic environment. At the protein scale, the scaffolding of multi-domain or multi-enzyme complexes that catalyze complex biosynthetic reactions inspires the design of protein or nucleic acid scaffolds that control the proximity of enzyme partners, increasing pathway flux. These strategies have been employed in a range of applications, from the remediation of toxic pollutants[20] to the production of biofuels[21], enabling unfavorable reactions and significantly boosting production. Given the importance of spatial regulation in natural systems, these strategies will likely be broadly applicable for synthetic biology and metabolic engineering.

Microbial assemblages

Like the nested symbiosis of methanogenic archaea inside *N. ovalis* in the cockroach gut, the close association of hydrogen producing protozoa and methanogens occurs in the gut of many other cellulose-digesting animal species, either episymbiotically in the cow rumen[22] or endosymbiotically in the termite hindgut[23]. This association allows for the complex metabolic breakdown pathways to be split between several members of the community while maintaining the spatial arrangement required for the efficient transfer of metabolic intermediates. Hydrogen transfer between different

organisms also occurs in other anaerobic environments where complex molecules are broken down by microbial consortia, such as putrefying sludge[9].

The inner life of sludge

Anaerobic wastewater sludge is home to highly structured microbial aggregates, sometimes millimeters across, that organize the metabolic reactions necessary for the complete breakdown of organic matter. These granules are arranged in three basic layers[9,24], each made up of multiple species, many of which remain uncharacterized and uncultured[25]. The aggregate is surrounded by an outer layer of acidogenic bacteria that break down biomass into organic acids. These acids are further broken down by the next layer of acetogenic, hydrogen-producing bacteria. In the center, multiple strains of methanogenic archaea including species of the genus *Methanosaeta*, *Methanobacterium*, *Methanospirillum*, and *Methanosarcina*[26] consume the hydrogen and carbon dioxide produced by the outer layers to generate methane[9] (Fig. 2A). The coupling of these metabolic reactions in granules ensures the efficient transfer of intermediates between the different cell types, improving the efficiency of biomass breakdown. In the case of propionate[27] or butyrate[28] degradation, the rapid consumption of hydrogen by the methanogenic core is required to drive the otherwise thermodynamically unfavorable reactions[9].

Intercellular spatial organization can also protect sensitive cells and enzymes from the external environment as well as help protect a cell from itself. Photosynthetic microorganisms have evolved a range of spatial organization schemes to isolate

metabolic enzymes that are inhibited[29] or damaged[30] by oxygen, a natural byproduct of photosystem II. In diazotrophic organisms, nitrogen fixation is performed by enzymes that are highly sensitive to atmospheric oxygen, and a number of strategies isolate the nitrogenase temporally or spatially[30]. In filamentous cyanobacteria such as *Anabaena* or *Nostoc* species, oxygen-sensitive nitrogenase is expressed only in terminally differentiated heterocyst cells that do not express photosystem II and are gas tight. These "multicellular" cyanobacteria exchange nutrients between the photosynthetic vegetative cells that can fix carbon and adjacent heterocysts that fix nitrogen, requiring a highly regulated system for patterning which cells differentiate to ensure proper distribution of nutrients along the filament[31] (Fig. 2B).

Nitrogen fixation by prokaryotes is crucial to the planetwide ecosystem, and is the basis for a variety of spatially organized symbioses. In addition to *Anabaena,* similar nitrogen-fixing filamentous bacteria are found in complex, highly structured endo- or episymbiotic relationships with plants, fungi, and even other bacteria such as *Hoeflea anabaenae* sp. nov., which can be found in epiphytic symbiosis attached to *Anabaena* heterocysts[32].

Given the power of ordered microbial communities, replicating these complex phenotypes has become a research priority. Tools to measure the spatially ordered metabolic interactions between different species in aggregates and biofilms[33], and engineer cell-cell communication[34] and programmed cell death[35], as well as advanced techniques such as microfluidics that allow for precise control over cell behavior have the potential to enable microbial "tissue engineering"[36], where microbial communities

are designed and spatially organized to produce or break down complex metabolites. Such multicellular microbial designs can be used not only to improve the transfer of pathway intermediates in metabolic reactions split across different species, but also to protect sensitive cells from harmful conditions. For example, oxygen-consuming microbes are frequently found in the outer layers of anaerobic aggregates, protecting the interior obligate anaerobes from oxygen damage[36]. Similarly, the outer layer of sludge granules can harbor microbes that degrade dangerous toxins, augmenting the overall fitness of the sludge while protecting the interior community from the toxins.

This protective strategy is exemplified by sludge granules tasked with degrading pentachlorophenol (PCP), a pesticide and disinfectant that is a common pollutant, contaminating soil and water. PCP inhibits the growth of most bacteria, but some species, like *Desulfitobacterium frappieri* PCP-1, can degrade the toxin. Supplementing anaerobic sludge reactors with *D. frappieri* PCP-1 protects the other species and allows for the efficient breakdown of PCP, with the exogenous bacteria colonizing the outside of the sludge granules[37]. Such methods are more difficult in the soil, where the presence of heavy metal co-pollutants such as mercury can harm PCP-degrading strains and inhibit bioremediation efforts. Here, a structured multi-species arrangement can thrive where individual strains cannot. *Ralstonia metallidurans* can efficiently reduce mercury and *Sphingobium chlorophenolicum* can degrade PCP, but in the presence of both mercury and PCP these strains are unable to survive alone or even in well-mixed co-culture. Biomimetic sludge granules, assembled via microfluidic laminar flow deposition,

demonstrated that wrapping *S. chlorophenolicum* in a layer of *R. metallidurans* allows for the efficient remediation of both mercury and PCP[20] (Fig. 2B).

Spatial structuring and aggregation of single-celled microbial species into multicellular assemblages can evolve quickly in directed evolution conditions. Cell aggregates reminiscent of those present in wastewater sludge were evolved in a co-culture of *Geobacter metallireducens* and *Geobacter sulfurreducens* grown on ethanol and fumarate. *G. metallireducens* is capable of consuming ethanol via Fe(III) reduction, but cannot use fumarate as an electron acceptor. Likewise, *G. sulfurreducens* uses fumarate as an electron acceptor but does not metabolize ethanol. After several months, cell aggregates appeared in the culture, significantly improving the time needed to consume the ethanol in the media[38] (Fig. 2C). These “great balls of evolution” facilitated the transfer of acetate and hydrogen from *G. metallireducens* consumption of ethanol to *G. sulfurreducens,* which could in turn reduce fumarate. Interestingly, these aggregates also formed when the hydrogenase genes in *G. metallireducens* were deleted, indicating that direct electron exchange could occur between the cell clusters in the aggregate. Microbial multicellularity takes many forms in nature[39] and can evolve in laboratory timescales.

Designing stable co-cultures

Such spatially controlled evolution and engineering of microbial consortia represents an alternative approach to synthetic biology, where the arrangement of different microbial species enables the design of a new biological function without

genetic engineering[20]. Minimal genetic alteration is required to set up a system of cross-feeding microorganisms, a first step for ecosystem-level design. Engineered co-culture based on auxotrophy was shown to lead to stable cooperation in *Saccharomyces cerevisiae*, with one strain auxotrophic for lysine but overproducing adenine and its partner auxotrophic for adenine but overproducing lysine[40]. Systems-level simulations of microbial cross-feeding in *Escherichia coli* identified significant metabolic synergy between auxotrophic strains without the need for additional engineering[41]. Based on these simulations, 46 conditional lethal *E. coli* auxotrophs were experimentally tested for their ability to grow in pairs. Of the 1035 pairs tested, 17% showed significant cooperative growth in minimal media, in agreement with the computational results. These experimentally validated cross-feeding pairs may form the basis for new engineered microbial consortia. Furthermore, this work demonstrated that network-scale metabolic modeling approaches commonly used for the design of monocultured species are also useful for the *a priori* design of stable co-cultures[41].

Synthetic cross-feeding can be useful for designing microbial consortia, as well as improving our understanding of microbial cooperation dynamics[11], but spatial engineering is often crucial for mixtures of organisms that do not normally coexist in nature to prevent one organism from dominating the culture. The importance of spatial control was demonstrated for an artificial “reciprocal syntrophy” designed between Nitrogen-fixing *Azotobacter vinelandii,* β-lactamase expressing *Bacillus licheniformis*, and cellulose-degrading *Paenibacillus curdlanolyticus.* These three strains can theoretically work together to grow in nitrogen-poor media containing penicillin and

carboxymethyl-cellulose as the only carbon source, but the community is unstable under well-mixed conditions. The cells could only grow cooperatively when they were physically isolated from one another in a microfluidic device that allowed for exchange of liquid media without the cells coming into direct contact[42]. Mathematical models of the ecological dynamics of such cooperating or competing[43] strains show that well-mixed populations are often less stable and less diverse than spatially ordered communities.

Designing complex consortia

The cell- and population-level modularity of microbial communities makes them dynamic and adaptable in the face of environmental change, and has great potential for synthetic biology applications in biosensing, bioremediation, chemical production, as well as future applications in health and medicine. As discussed previously, many complex microbial communities are found in animal digestive tracts; the human gut itself is home to hundreds of microbial species[44]. The successful application of fecal transplantation to patients recovering from *Clostridium difficile* infections suggests that engineering the species composition of the human microbiome is a valuable therapeutic strategy[45].

The value of microbial consortia in synthetic biology[13] and metabolic engineering[46] has been explored in many contexts. At the same time, study of microbial communities through metagenomics[47] has made significant a1dvances in our understanding of microbial ecology in the soil[48] and the human body[49], among many others. These DNA sequence-based approaches will undoubtedly improve our ability to

engineer microbial consortia, but rarely provide insight into the small-scale spatial organization that can have a large impact on the metabolic function of the community. In synthetic biology, microfluidics approaches can demonstrate the principles of spatial organization and community dynamics but they lack the versatility and robustness of natural microbial consortia. Approaches that take advantage of the natural abilities of microbes to form assemblages, editing the community by addition of specific species[37] or evolving new behaviors[38] are currently best able to generate engineered systems. As our understanding of community dynamics and our ability to engineer cell-cell communication[50] and self-assembling cell scaffolds[51] improves we foresee the design of ordered multispecies communities with broad applications in biological engineering.

Organelles and Microcompartments

Compartmentation within cells offers many of the advantages of microbial assemblages—metabolic reactions can be isolated within a compartment, and concentration gradients can be exploited between compartments. Eukaryotic evolution is marked by the transition from an endosymbiotic microbial assemblage to the fully integrated organelles in modern cells[52,53]. Likewise, the vertically transmitted archaea that reside within *N. ovalis* can be thought of as “second-generation” organelles: they are formerly free-living archaea attached to hydrogenosomes, which are formerly free-living bacteria. The continuing evolution of intercellular compartmentation is a compelling inspiration for biological design.

Eukaryotic organelles

Intracellular spatial organization of cellular functions by membrane bound organelles is one of the defining features of eukaryotic cells. In addition to the aforementioned hydrogenosome, other organelles such as peroxisomes, vacuoles, mitochondria, and chloroplasts isolate specialized metabolic reactions from the cytosol. The properties of these organelles can be modified or mimicked to improve engineered pathways.

Methylotrophic yeasts are capable of utilizing methanol as a sole carbon source, which is an impressive metabolic feat considering that hydrogen peroxide and formaldehyde are necessary intermediates for methanol metabolism. Yeasts such as *Pichia pastoris*, *Hansenula polymorpha*, and others are capable of consuming methanol via morphological adaptation of their peroxisomes, membrane-bound organelles that perform a variety of metabolic tasks[54,55]. When these yeasts are grown on methanol, peroxisome biogenesis is upregulated such that peroxisomes become the dominant structural feature of the cell[56,57] (Fig. 3A).

Several enzymes within the peroxisome produce and destroy toxic intermediates during methanol metabolism. Alcohol oxidase (AOX) catalyzes the conversion of methanol to formaldehyde, releasing hydrogen peroxide. Dihydroxyacetone synthase takes a five-carbon sugar, xylulose-5-phosphate, and reacts it with one-carbon formaldehyde to produce the three-carbon compounds dihydroxyacetone and glyceraldehyde-3-phosphate. Catalase enzymes convert the toxic hydrogen peroxide into water and oxygen. The action of these three enzymes together must occur in the

peroxisome and mis-targeting any of them prevents growth on methanol[54]. AOX is vastly more abundant than other peroxisome enzymes when methanol is present. AOX can be tightly packed to form paracrystalline structures which fill nearly the entire peroxisome. Catalase and dihydroxyacetone synthase are believed to be trapped within the AOX superstructure[58]. This structural arrangement may serve to limit buildup of formaldehyde and hydrogen peroxide while maintaining high pathway flux.

Organelle engineering

Targeting of proteins to the peroxisome can be achieved by addition of a number of short targeting sequences[59] for the isolation of any desired metabolic pathway. In a similar synthetic approach, N-terminal targeting sequences were used to place methyl halide production pathways in the *S. cerevisiae* vacuole, where the majority of S-adenosyl methionine (SAM), a required cofactor, is sequestered[60]. Although this intervention resulted in a less than 25% increase in methyl halide production rates, the addition of methionine to the media along with vacuole targeting increased methyl halide production rates by over 400%. Methionine induces the overproduction of SAM in yeast[61], allowing SAM to accumulate in the vacuole for methyl halide synthesis[60]. As with AOX production in peroxisomes, this work underscores the context dependence of organellar metabolic activity; the biogenesis and metabolic activity of organelles are highly regulated.

Another emerging strategy for organelle generation is induced endosymbiosis. Mitochondria, hydrogenosomes, and chloroplasts all appear to have evolved from

ancient free-living prokaryotes (α-proteobacteria for mitochondria and hydrogenosomes, cyanobacteria for chloroplasts) that became endosymbionts within eukaryotic cells[52,53]. In an attempt to recapitulate the function of chloroplasts in non-photosynthetic cell types, researchers introduced the cyanobacterium *Synechococcus elongatus* PCC 7942 into the cytoplasm of zebrafish embryos, Chinese Hamster Ovary (CHO) cells, and human macrophages[62]. This was achieved via manual injection, invasin-mediated cell invasion, and phagocytosis, respectively. Intriguingly, the cyanobacteria were well-tolerated by the cells in each case—*E. coli* cells in contrast kill zebrafish embryos post-injection—indicating that cyanobacteria may be engineered to provide useful products to a host cell endosymbiotically. An important next step will be to determine how to appropriately construct mutual metabolic exchanges between a synthetic endosymbiont and the host cell, such that selection pressures favor maintaining endosymbiosis. Since endosymbiotic metabolic exchanges are similar to cooperative metabolic exchanges between co-cultures, modeling approaches developed for co-culture design may also facilitate the design of endosymbioses[41].

Prokaryotic microcompartments

Compartmentation was once thought to be a hallmark of eukaryotic systems, but the discovery of bacterial microcompartments (BMCs) has demonstrated that spatial organization matters at the micron scale. BMCs are typically proteinaceous shells of defined geometric shapes that resemble viral capsids, although structural studies have shown that the characteristic BMC fold is unrelated to known viral protein structures[6].

Interestingly, BMCs have been identified in a wide range of roles, from carbon fixation to the metabolism of toxic compounds[6] (Fig. 3B).

The carboxysome is a particularly well-studied BMC involved in carbon fixation. Carboxysomes house rubisco, the key Calvin cycle enzyme that fixes gaseous carbon dioxide to ribulose-1,5-biphosphate to form 3-phosphoglycerate. All known rubisco enzymes suffer from an inability to distinguish between carbon dioxide and oxygen; both cyanobacteria and higher plants address this by expressing large amounts of rubisco and by spatially sequestering it from the oxygenic photosynthetic machinery[6,63]. As with AOX expression in the peroxisome, concentration of rubisco in the carboxysome offers sufficient protection to allow both photosynthesis and carbon fixation to coexist in cells that are essentially the same size as *E. coli*. As with *Anabaena* heterocyst formation (Fig. 2B), the spacing of carboxysomes within cyanobacterial cells is highly structured to ensure even distribution of carbon fixation machinery upon cell division[63] (Fig. 3B).

The carboxysome appears to be more than a passive protein container. The two enzymes in carboxysomes, rubisco and carbonic anhydrase, are differentially localized. Rubisco is packed tightly into the carboxysome interior[64], while carbonic anhydrase, which converts aqueous bicarbonate to gaseous carbon dioxide, is bound to the interior of the protein shell[65]. Establishing a concentration differential of bicarbonate across the shell boundary appears to be essential, as expressing carbonic anhydrase in the cytosol disrupts carbon fixation by carboxysomes[66]. How cells constrain enzyme function to within intact BMC when proteins must be translated in the cytosol remains unknown.

There is mounting evidence that BMCs regulate the transport of metabolites across the BMC shell. Particularly intriguing are pore proteins such as CsoS1D. CsoS1D trimers appear to be able to adopt two conformations, one with an open central pore and another in which the pore is shut[67]. CsoS1D is important to carboxysome structure, and improves the function of heterologously expressed carboxysomes[17]. Other BMCs, such as 1,2-propanediol utilization (PDU) metabolosomes, may be capable of transporting large molecules such as cobalamin (vitamin B12) in and out of intact BMCs[68,79]. Recent efforts have successfully targeted proteins to BMCs in *Salmonella*[70] as well as expressed functional PDU and carboxysome BMCs in *E. coli*[17,71].

Custom heterologous microcompartments with targeted enzyme pathways could perform a wide range of tasks, enabling reactions that would be otherwise unfavorable inside a bacterial cell. Current knowledge of protein targeting and BMC biogenesis has only just scratched the surface of the potential toolset these compartments will provide for metabolic engineering. Given the diverse BMC capabilities that have already been discovered, further study of BMC structures, pore complexes, and targeting sequences is likely to allow BMCs to be repurposed for new metabolic pathways in industrial strains.

Enzymes and scaffolds

The same principles that organize the spatial optimization of metabolic pathways across microbial communities and inside cellular compartments apply to the scale of

metabolic enzymes. Prokaryotic cells were once thought of as simple bags of enzymes, with random diffusion controlling the motion and position of enzymes and reactants in the cytoplasm, but researchers have since observed spatial regulation in even the simplest cells. By isolating pathways in this way, metabolic intermediates are more likely to go from one enzyme to the next in the pathway without being diverted to other reactions, boosting the flux through the reaction, as well as protecting unstable intermediates from degradation or protecting the cell in from the buildup of any toxic intermediates.

The cellulosome

For synthetic biologists, the structural intricacy of cellulose is both an inspiration for designing new biomaterials and a technical barrier for producing cellulosic biofuels. Cellulose in plants is produced by a 36-enzyme “rosette” that deposits 36 polysaccharide chains that entwine to form a cellulose elementary fibril[72]. These fibrils are woven along with hemicellulose, lignin, pectin, and other associated proteins to form a dense crystalline matrix that is notoriously difficult to break down[73].

As a consequence, spatial organization plays an important role in cellulose breakdown not only at the cellular scale in the arrangements of protozoa and archaea, but also at the protein scale in the organization of the enzymes involved in the digestion of the cellulose polymers. One of only a handful of enzymatic systems capable of breaking down cellulose, the cellulosome is a megadalton-scale extracellular complex on the surface of anaerobic bacteria such as *Clostridium thermocellum*. Central to this

20

complex is a non-catalytic scaffold protein that is decorated with modular scaffoldin domains and a carbohydrate binding domain. Cellulosome enzymes possess dockerin domains, which bind to the scaffoldin domains on the primary scaffold. The scaffold performs two key functions: it co-localizes cellulases, hemicellulases, and other enzymes, and it also directly attaches the bacterial cell to the plant cell wall. In this manner, the cellulosome is sandwiched between the plant matter and the host cell, giving the host preferential access to the degradation products[74,75] (Fig. 4A).

The remarkable modularity of cellulosomes has led to the construction of synthetic cellulosomes. Functional small-scale cellulosomes have been expressed heterologously in *Clostridium acetobutylicum*[76], *Saccharomyces cerevisiae*[14,77], and *Bacillus subtilis*[78]. In addition, novel cellulosomes have been constructed by tagging cellulase and xylanase enzymes that are normally not scaffold-bound with dockerin domains[79]. Enzymes can be targeted to specific sites on synthetic cellulosomes by using a range of divergent scaffoldin-dockerin pairs[80]; these constructs may prove useful for applications beyond cellulose degradation.

Protein-scale assembly lines

Enzyme complexes organize metabolic reactions that produce complex metabolites and polymers in similarly modular fashion inside of the cell. Multifunctional enzymes made up from several distinct protein domains commonly bring together active sites that share reactants, thus limiting the range an intermediate must diffuse before the next enzymatic conversion. A small number of enzymes further reduce the distance

an intermediate must travel between active sites with substrate tunnels through which unstable compounds can travel one dimensionally between domains[16]. These complex enzyme structures are remarkable; for example, carbamoyl phosphate synthase channels ammonia from the hydrolysis of glutamine for nearly 100Å through several protein domains. Each of these multifunctional, substrate tunneling enzymes are structurally distinct, evolving independently[16]. These bespoke structures are effective at solving the diffusion problem for individual reactions and pathways, but the specificity of the enzymes makes recombination of the domains and "reprogramming" of new multifunctional enzyme chemistry unlikely.

On the other hand, polyketide synthases are another class of multifunctional enzyme complexes that have enjoyed special attention from bioengineers due to their evolutionary flexibility. Polyketides are a large and diverse group of complex natural products with many important medicinal functions, including antibiotic and antitumor activity[18]. Like other evolutionarily related iterative chain-lengthening complexes such as fatty acid synthase[81] or non-ribosomal peptide synthase[82], polyketide synthases act as an "assembly line," passing the growing chain from one domain to the next (Fig. 4B). "Unnatural" natural products can be produced through the synthetic rearrangement of these enzyme domains in the cell[83-85].

Programming multifunctional enzyme complexes to produce specific compounds remains a significant challenge despite decades of effort to understand the function and specificity of each enzyme domain. Efficient transfer of the growing acyl chain between recombined polyketide synthase domain depends on the substrate specificity within

modules[15], as well as the sequence of the polypeptide linker between modules[86]. Analysis of protein-protein interactions[82] and sequence co-evolution[19] provides important information to aid the design of novel enzyme complexes.

Similar analysis of protein-protein interaction co-evolution has allowed for the "rewiring" of bacterial two-component signal transduction[87]. Indeed, signal transduction provides valuable analogies for the spatial organization of metabolism. In particular, scaffolding of promiscuous signaling domains ensures the proper interactions and function of eukaryotic signaling cascades, and rearrangement of scaffold domain interactions can alter pathway function[88-90]. Linking metabolic enzymes to protein-protein interaction domains derived from eukaryotic cell signaling can likewise scaffold synthetic metabolic pathways, preventing the diffusion of toxic intermediates, improving the stoichiometric ratios of enzymes, and significantly boosting pathway flux[8,9].

Electron Transfer

Physical interaction between metabolic enzymes can boost flux through many metabolic pathways, but it is absolutely crucial in the case of electron-transfer reactions, where electrons must quantum-mechanically tunnel between electron-binding metal clusters at an optimal distance of 14Å[91]. Nowhere is this more apparent than in the membrane bound electron transfer chains of the mitochondria or the chloroplast (Fig. 5A), as well as the electrogenic pathways of bacteria such as *Shewanella* or *Geobacter*. Electrogenic bacteria utilize membrane-bound electron transport chains to "breathe" extracellular metals, dumping excess reducing power through their membranes and

generating electrical current[92]. In each of these electron transport systems, linked cytochromes, quinones and other iron-sulfur cluster containing oxido-reductases are anchored in the membrane; this decreases the diffusional difficulty of finding the proper interacting partner to only two dimensions.

In cytoplasmic enzymes, recombination of ferredoxin domains creates a currency of electron-transferring potential. Multifunctional electron-transfer enzymes can be designed through the recombination and fusion of multiple such domains, connecting electron transfer enzymes from multiple organisms into large protein fusions that function in vitro or in vivo. These include a number of fusions between cytochrome P450, an important class of monooxygenase enzymes involved in a large number of metabolic reactions[93]. For example, mammalian cytochrome P450 has been fused to yeast[94] or bacterial[95] reductases. Fusions have also been designed between ferredoxins and their partner ferredoxin-oxidoreductases[96], including [Fe-Fe]-hydrogenase enzymes, which use electrons carried by ferredoxin to reduce protons to hydrogen gas. Fusing hydrogenases to compatible ferredoxins boosted production of hydrogen through a synthetic pathway nearly 5-fold[97]. In vitro fusion of hydrogenase directly to photosystem I has allowed for efficient photobiological hydrogen production[98,99](Fig. 5B).

Tethering of electron transfer enzymes to artificial scaffold proteins also boosts hydrogen production *in vivo* from a synthetic electron transfer pathway by up to 3 fold[97]. Self-assembling scaffolds made from RNA strands that fold into one dimensional ribbons or two-dimensional sheets can also be designed to tether electron transfer

pathways, boosting hydrogen production from a heterologous pathway by nearly 50 fold[100] (Fig. 5C). RNA scaffolds are orthologous systems that can be modified to function in many systems, creating a generalizable tool for the spatial organization of many metabolic pathways.

Emerging synthetic scaffold approaches

To date, synthetic scaffolds based on protein[8], RNA[100], and DNA[101] have been demonstrated to improve the performance of metabolic pathways. This variety of scaffold types will allow engineers to choose scaffolds with properties that best complement their target metabolic pathway. For example, RNA scaffolds have been demonstrated to self assemble into many 100 nm scale complexes that dominate the interior of the cell[100]. In contrast, DNA scaffolds permit precise control over the number of enzymes bound to the scaffold, based on the number of binding sites in the genome[101]. All scaffolds offer increased control over pathway stoichiometry and enzyme proximity. Such methods can be generally applied to a wide range of metabolic pathways, but are limited by the in vivo dynamics of large heterologous protein complexes, including aggregation. Improvements in the fine-tuning of scaffold expression and targeting will improve the function of these useful structures.

Conclusions

Spatial optimization of metabolism occurs at many scales that can potentially be exploited in synthetic systems. Multicellular complexes separate incompatible chemical

reactions, and couple reactions that would be thermodynamically unfavorable in isolation. Within cells, compartments such as the carboxysome, vacuole, or peroxisome concentrate reactions and pathways that need special cellular conditions to function properly and may be toxic to cytoplasmic contents. At the protein scale, the substrate channeling of multifunctional enzymes and the scaffolding of multi-domain or multi-enzyme complexes enhance complex biosynthetic reactions such as fatty acid synthesis, polyketide synthesis, and non-ribosomal protein synthesis, as well as difficult breakdown reactions such as cellulose degradation.

We have presented several early efforts to translate our understanding of spatial organization at various scales into engineered biological systems that incorporate these principles. To improve upon spatial engineering approaches, we must consider the issues that currently complicate biological design at each scale. At the level of cells and communities, while over a century of microbiology research has focused on growing bacteria in monocultures, the design of stable co-cultures still represents a significant challenge. An intermediate approach can be to couple an engineered species with a wild-type species that performs a singular task. For example, *Actinotalea fermentans* is a bacterium that ferments cellulose to ethanol and acetate. The methyl halide producing yeast discussed previously were found to form a stable co-culture with *A. fermentans*, producing methyl halides from cellulosic feedstocks[54].

Likewise, inside the cell, approaches that interface natural spatial organization mechanisms with engineered functions are best able to take advantage of the benefits of three dimensional patterning in metabolic pathways. In eukaryotes, targeting of

proteins to organelles[60] using well-understood mechanisms[59] can help boost the function of synthetic metabolic pathways. Synthetic reconstitution of microcompartments in prokaryotes[17] can improve our understanding of these remarkable structures, taking us closer to being able to design custom prokaryotic compartments. Spatial organization at the protein scale is better understood, and has seen the most progress in the design of recombinant protein docking and scaffolding[8,14,84]. These approaches have significantly expanded the metabolic toolkit into three dimensions and will likely play an important role in a range of future pathway designs.

Innovation in intracellular engineering will require the characterization of many new biological parts—such as organelle targeting sequences, transmembrane transporters, and BMC pores—parts that have evolved to support natural metabolic exchanges. Integrating these new parts into engineered metabolic pathways comes with an increased cost to the researcher, in that it increases the combinatorial complexity of pathway designs that must be tested. In today's trial and error phase of synthetic biology, this cost is not trivial. However, given the observed benefits of spatial organization in evolved systems, spatial approaches are likely to be essential to solving the most difficult biological design challenges.

Across all scales, biological form is defined by function; many historical milestones in biological research have recognized this property in new contexts. We have presented examples of spatial contexts that are important or even essential to metabolic function. The canonical metabolic network diagram reveals the vast array of metabolic conversions that have evolved in biological systems, but hides the spatial

contexts that make many of these conversions possible. As synthetic biologists, we seek to engineer metabolic networks; by understanding and reengineering natural systems we can extend the metabolic network diagram into uncharted territory.

28

Acknowledgments

The authors would like to thank Johaness Hackstein, Zarath Summers, Derek Lovley, David Savage, and Bruno Afonso for the use of images. PMB is supported by fellowships from the Harvard University Center for the Environment and the NSF Synthetic Biology Engineering Research Center. PAS acknowledges support from the Radcliffe Institute of Advanced Study, the Wyss Institute for Biologically Inspired Engineering, the DOD Department of the Army, and ARPA-E DOE.

Competing Interests Statement

The authors declare no competing financial interests.

Figure Legends

Figure 1. Nested endosymbiosis in the cockroach hindgut. A.) Electron micrograph of the eukaryotic ciliate *Nyctotherous ovalis*, an important symbiont for cockroach cellulose digestion[3]. *N. ovalis* produces ATP anaerobically in an organelle termed the hydrogenosome, via a process that requires the continual production of hydrogen. Hydrogenosomes (H) are surrounded by methanogenic endosymbiotic *Methanobrevibacter*-like archaea (black dots) that consume hydrogen and ensure that the ATP producing reactions remain thermodynamically favorable. N, macronucleus; n, micronucleus; V, vacuole. (Image courtesy of Johannes Hackstein, Nijmegen, The Netherlands).

Figure 2. Microbial assemblages can perform coupled metabolic reactions. A.) Anaerobic sludge granules are structured aggregates of many species of microorganisms that can be millimeters across, and are composed of multiple layers[26]. The granule is surrounded by acidogens that break down complex organic molecules into acids. These acids are consumed by acetogens that form a middle, hydrogen producing layer. Hydrogen and carbon dioxide produced by these outer layers is consumed by the central core methanogens[7]. B.) Cellular differentiation in nitrogen starved *Anabaena* filaments is highly regulated to ensure proper spacing of oxygen-sensitive nitrogen-fixing heterocysts and oxygen-producing, photosynthetic carbon-fixing vegetative cells[30,31]. C.) A bilayered microbial fiber assembled with microfluidics is able to break down both mercury (Hg) and PCP. Mercury-reducing *Ralstonia metallidurans* surrounds the PCP-degrading species *Sphingobium chlorophenolicum*, reducing mercury and protecting the inner layer, which is free to degrade PCP. Well-mixed cultures of these same species are unable to degrade PCP, as *S. chlorophenolicum* is sensitive to mercury[20]. D.) Aggregates of the closely related species *Geobacter metallireducens* and *Geobacter sulfurreducens* form when the co-culture must consume ethanol using fumarate as an electron donor. Neither species is capable of growth on ethanol and fumarate in isolation. Initial co-cultures of these two species did not form aggregates and grew poorly on ethanol. Directed evolution yielded aggregating mutants with vastly improved growth rates; aggregation facilitates direct electron transfer between the two species[38].

Figure 3. Compartmentalization isolates metabolic pathways in eukaryotic and prokaryotic cells. A.) Yeasts that can metabolize methanol isolate methanol breakdown to the peroxisome, protecting the cell from the release of formaldehyde and hydrogen peroxide, the products of methanol metabolism[56]. AOX: Alcohol oxidase, DHS: Dihydroxyacetone synthase, CAT: catalase B.) The carboxysome is a protein-shelled microcompartment that concentrates the carbon fixation machinery in cyanobacteria[17,63]. Carboxysomes are further ordered along the cell axis. This ensures equal distribution upon cell division, and may play a role in maintaining optimal carbon fixation conditions[63]. (Images courtesy of David Savage, Berkeley, USA and Bruno Afonso, Boston, USA)

Figure 4. Enzyme complexes aid in the breakdown and production of complex molecules. A.) The cellulosome is a modular complex for degrading cellulose. Carbohydrate binding modules (CBM) attach the cellulosome to cellulose. Dockerin/cohesin interactions allow specific pairings of enzymes to the scaffold, and also attach the entire complex to the host cell. Designer cellulosomes utilize combinations of different cohesin/dockerin pairs, allowing the attachment of multiple distinct proteins to the same scaffold[79]. B.) The erythromycin polyketide synthase (PKS) 6-deoxyerythronolide B synthase (DEBS) is one of the best studied PKS complexes. It is composed of three multifunctional enzymes each made up of multiple repeated

modules and subdomains that transfer the growing acyl chain. Recombining the position of these domains can yield novel polyketides[84].

Figure 5. Electron transfer pathways are optimized through spatial organization. A.) Transfer of electrons (dotted line) through the photosynthetic machinery scaffolded in the chloroplast thylakoid membrane. Photons excite electrons at PSII and PSI. Water splitting in the lumen and NADP+ reduction are spatially separated by the thylakoid membrane. PSII: photosystem II, PQ: plastoquinone, cyt: cytochrome, PSI: photosystem I, Fd: ferredoxin, FNR: ferredoxin:nadph oxidoreductase. B.) Direct linkage of photosystem I with the hydrogenase allows for *in vitro* transfer of electrons generated from sunlight to hydrogen gas[98,99] C.) Linking of the hydrogenase enzyme and ferredoxin to an RNA scaffold improves electron transfer efficiency and hydrogen production by more than 40 fold[100]. RNA aptamer domains specific for protein tags allow precise positioning of enzymes on the scaffold.

References

1. Gijzen, H. J. & Barugahare, M. Contribution of anaerobic protozoa and methanogens to hindgut metabolic activities of the American cockroach, *Periplaneta americana*. *Appl Environ Microbiol* **58**, 2565–2570 (1992).
2. van Hoek, A. H. *et al.* Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. *Mol Biol Evol* **17**, 251–258 (2000).
3. Akhmanova, A. *et al.* A hydrogenosome with a genome. *Nature* **396**, 527–528 (1998).
4. Ushida, K. Symbiotic methanogens and rumen ciliates. *(Endo) symbiotic Methanogenic Archaea* 25–34 (2011).
5. Boxma, B. *et al.* An anaerobic mitochondrion that produces hydrogen. *Nature* **434**, 74–79 (2005).
6. Yeates, T. O., Crowley, C. S. & Tanaka, S. Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution. *Annual review of biophysics* **39**, 185–205 (2010).

7. MacLeod, F., Guiot, S. & Costerton, J. Layered structure of bacterial aggregates produced in an upflow anaerobic sludge bed and filter reactor. *Appl Environ Microbiol* **56**, 1598–1607 (1990).
8. Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27**, 753–759 (2009).
9. Lee, H., DeLoache, W. C. & Dueber, J. E. Spatial organization of enzymes for metabolic engineering. *Metab Eng* (2011).doi:10.1016/j.ymben.2011.09.003
10. Solomon, K. V. & Prather, K. L. J. The zero-sum game of pathway optimization: Emerging paradigms for tuning gene expression. *Biotechnol J* **6**, 1064–1070 (2011).
11. Brenner, K., You, L. & Arnold, F. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* **26**, 483–489 (2008).
12. Zuroff, T. R. & Curtis, W. R. Developing symbiotic consortia for lignocellulosic biofuel production. *Appl Microbiol Biotechnol* (2012)
13. Shin, H.-D., McClendon, S., Vo, T. & Chen, R. R. *Escherichia coli* binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Appl Environ Microbiol* **76**, 8150–8159 (2010).
14. Tsai, S. L., Oh, J., Singh, S., Chen, R. & Chen, W. Functional Assembly of Minicellulosomes on the *Saccharomyces cerevisiae* Cell Surface for Cellulose Hydrolysis and Ethanol Production. *Appl Environ Microbiol* **75**, 6087–6093 (2009).
15. Watanabe, K. Understanding Substrate Specificity of Polyketide Synthase Modules by Generating Hybrid Multimodular Synthases. *Journal of Biological Chemistry* **278**, 42020–42026 (2003).
16. Weeks, A., Lund, L. & Raushel, F. M. Tunneling of intermediates in enzyme-catalyzed reactions. *Current Opinion in Chemical Biology* **10**, 465–472 (2006).
17. Bonacci, W. *et al.* Modularity of a carbon-fixing protein organelle. *Proceedings of the National Academy of Sciences* **109**, 478–483 (2012).
18. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
19. Thattai, M., Burak, Y. & Shraiman, B. I. The origins of specificity in polyketide synthase protein interactions. *PLoS Comput Biol* **3**, 1827–1835 (2007).
20. Kim, H. J., Du, W. & Ismagilov, R. F. Complex function by design using spatially pre-structured synthetic microbial communities: degradation of pentachlorophenol in the presence of Hg(ii). *Integr. Biol.* **3**, 126 (2011).
21. Goyal, G., Tsai, S.-L., Madan, B., DaSilva, N. A. & Chen, W. Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. *Microb Cell Fact* **10**, 89 (2011).
22. Vogels, G. D., Hoppe, W. F. & Stumm, C. K. Association of methanogenic bacteria with rumen ciliates. *Appl Environ Microbiol* **40**, 608–612 (1980).

23. Lee, M. J., Schreurs, P. J., Messer, A. C. & Zinder, S. H. Association of methanogenic bacteria with flagellated protozoa from a termite hindgut. *Curr Microbiol* **15**, 337–341 (1987).
24. Gonzalez-Gil, G. *et al.* Cluster structure of anaerobic aggregates of an expanded granular sludge bed reactor. *Appl Environ Microbiol* **67**, 3683–3692 (2001).
25. Liu, W. T., Chan, O. C. & Fang, H. H. P. Characterization of microbial community in granular sludge treating brewery wastewater. *Water research* **36**, 1767–1775 (2002).
26. Sekiguchi, Y., Kamagata, Y., Nakamura, K., Ohashi, A. & Harada, H. Fluorescence in situ hybridization using 16S rRNA-targeted oligonucleotides reveals localization of methanogens and selected uncultured bacteria in mesophilic and thermophilic sludge granules. *Appl Environ Microbiol* **65**, 1280–1288 (1999).
27. Boone, D. R. & Bryant, M. P. Propionate-Degrading Bacterium, *Syntrophobacter wolinii* sp. nov. gen. nov., from Methanogenic Ecosystems. *Appl Environ Microbiol* **40**, 626–632 (1980).
28. Dwyer, D. F., Weeg-Aerssens, E., Shelton, D. R. & Tiedje, J. M. Bioenergetic conditions of butyrate metabolism by a syntrophic, anaerobic bacterium in coculture with hydrogen-oxidizing methanogenic and sulfidogenic bacteria. *Appl Environ Microbiol* **54**, 1354–1359 (1988).
29. Raven, J. A., Cockell, C. S. & La Rocha, De, C. L. The evolution of inorganic carbon concentrating mechanisms in photosynthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 2641–2650 (2008).
30. Kumar, K., Mella-Herrera, R. A. & Golden, J. W. Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol* **2**, a000315 (2010).
31. Meeks, J. C. & Elhai, J. Regulation of Cellular Differentiation in Filamentous Cyanobacteria in Free-Living and Plant-Associated Symbiotic Growth States. *Microbiol Mol Biol Rev* **66**, 94–121 (2002).
32. Stevenson, B. S. *et al.* Hoeflea anabaenae sp. nov., an epiphytic symbiont that attaches to the heterocysts of a strain of Anabaena. *Int. J. Syst. Evol. Microbiol.* **61**, 2439–2444 (2011).
33. Phelan, V. V., Liu, W.-T., Pogliano, K. & Dorrestein, P. C. Microbial metabolic exchange—the chemotype-to-phenotype link. *Nat Chem Biol* **8**, 26–35 (2011).
34. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
35. You, L., Cox, R. S., Weiss, R. & Arnold, F. H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428**, 868–871 (2004).

36. Markx, G. H., Andrews, J. S. & Mason, V. P. Towards microbial tissue engineering? *Trends Biotechnol* **22**, 417–422 (2004).
37. Lanthier, M., Tartakovsky, B., Villemur, R., DeLuca, G. & Guiot, S. R. Microstructure of Anaerobic Granules Bioaugmented with *Desulfitobacterium frappieri* PCP-1. *Appl Environ Microbiol* **68**, 4035–4043 (2002).
38. Summers, Z. M. *et al.* Direct Exchange of Electrons Within Aggregates of an Evolved Syntrophic Coculture of Anaerobic Bacteria. *Science* **330**, 1413–1415 (2010).
39. Shapiro, J. A. Thinking about bacterial populations as multicellular organisms. *Annu Rev Microbiol* **52**, 81–104 (1998).
40. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc Natl Acad Sci USA* **104**, 1877–1882 (2007).
41. Wintermute, E. H. & Silver, P. A. Emergent cooperation in microbial metabolism. *Mol Syst Biol* **6**, 1–7 (2010).
42. Kim, H. J., Boedicker, J. Q., Choi, J. W. & Ismagilov, R. F. Defined spatial structure stabilizes a synthetic multispecies bacterial community. *Proc Natl Acad Sci USA* **105**, 18188–18193 (2008).
43. Kerr, B., Riley, M. & Feldman, M. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* (2002).
44. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
45. Tvede, M. & Rask-Madsen, J. Bacteriotherapy for chronic relapsing *Clostridium difficile* diarrhoea in six patients. *Lancet* **1**, 1156–1160 (1989).
46. Shong, J., Diaz, M. R. J. & Collins, C. H. Towards synthetic microbial consortia for bioprocessing. *Curr Opin Biotechnol* 1–5 (2012).
47. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**, 525–552 (2004).
48. Daniel, R. The metagenomics of soil. *Nat Rev Micro* **3**, 470–478 (2005).
49. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* **449**, 804–810 (2007).
50. Choudhary, S. & Schmidt-Dannert, C. Applications of quorum sensing in biotechnology. *Appl Microbiol Biotechnol* 1–13 (2010).
51. Aldaye, F. A., Senapedis, W. T., Silver, P. A. & Way, J. C. A Structurally Tunable DNA-Based Extracellular Matrix. *J Am Chem Soc* **132**, 14727–14729 (2010).
52. Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
53. Howe, C. J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J. & Larkum, A. W. D. The origin of plastids. *Philos Trans R Soc Lond, B, Biol Sci* **363**, 2675–2685 (2008).

54. van der Klei, I. J., Yurimoto, H., Sakai, Y. & Veenhuis, M. The significance of peroxisomes in methanol metabolism in methylotrophic yeast. *Biochim Biophys Acta* **1763**, 1453–1462 (2006).
55. Yurimoto, H., Oku, M. & Sakai, Y. Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. *Int J Microbiol* **2011**, 101298 (2011).
56. van der Klei, I. J., Harder, W. & Veenhuis, M. Biosynthesis and assembly of alcohol oxidase, a peroxisomal matrix protein in methylotrophic yeasts: a review. *Yeast* **7**, 195–209 (1991).
57. Ozimek, P., Veenhuis, M. & van der Klei, I. J. Alcohol oxidase: a complex peroxisomal, oligomeric flavoprotein. *FEMS Yeast Res* **5**, 975–983 (2005).
58. Vonck, J. & van Bruggen, E. F. Architecture of peroxisomal alcohol oxidase crystals from the methylotrophic yeast Hansenula polymorpha as deduced by electron microscopy. *J Bacteriol* **174**, 5391–5399 (1992)
59. Roggenkamp, R. Targeting signals for protein import into peroxisomes. *Cell Biochem. Funct.* **10**, 193–199 (1992).
60. Bayer, T. S. *et al.* Synthesis of methyl halides from biomass using engineered microbes. *J Am Chem Soc* **131**, 6508–6515 (2009).
61. Lin, J.-P. *et al.* An effective strategy for the co-production of S-adenosyl-l-methionine and glutathione by fed-batch fermentation. *Biochemical Engineering Journal* **21**, 19–25 (2004).
62. Agapakis, C. M. *et al.* Towards a synthetic chloroplast. *PLoS ONE* **6**, e18877 (2011).
63. Savage, D. F., Afonso, B., Chen, A. H. & Silver, P. A. Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science* **327**, 1258–1261 (2010).
64. Iancu, C. V. *et al.* The Structure of Isolated Synechococcus Strain WH8102 Carboxysomes as Revealed by Electron Cryotomography. *J Mol Biol* **372**, 764–773 (2007).
65. So, A. K.-C. *et al.* A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. *J Bacteriol* **186**, 623–630 (2004).
66. Price, G. D. & Badger, M. R. Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO(2)-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO(2) Concentrating Mechanism. *Plant Physiol* **91**, 505–513 (1989).
67. Klein, M. G. *et al.* Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport. *J Mol Biol* **392**, 319–333 (2009).
68. Cheng, S. & Bobik, T. A. Characterization of the PduS Cobalamin Reductase of Salmonella enterica and Its Role in the Pdu Microcompartment. *J Bacteriol* **192**, 5071–5080 (2010).

69. Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. & Bobik, T. A. Bacterial microcompartments: their properties and paradoxes. *BioEssays* **30**, 1084–1095 (2008).
70. Fan, C. *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences* **107**, 7509–7514 (2010).
71. Parsons, J. B. *et al.* Biochemical and structural insights into bacterial organelle form and biogenesis. *J Biol Chem* **283**, 14366–14375 (2008).
72. Doblin, M. S., Kurek, I., Jacob-Wilk, D. & Delmer, D. P. Cellulose biosynthesis in plants: from genes to rosettes. *Plant and Cell Physiology* **43**, 1407–1420 (2002).
73. Ding, S.-Y. & Himmel, M. E. The Maize Primary Cell Wall Microfibril: A New Model Derived from Direct Visualization. *Journal of Agricultural and Food Chemistry* **54**, 597–606 (2006).
74. Fontes, C. M. G. A. & Gilbert, H. J. Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. *Annu Rev Biochem* **79**, 655–681 (2010).
75. Fischbach, M. & Voigt, C. A. Prokaryotic gene clusters: A rich toolbox for synthetic biology. *Biotechnol J* **5**, 1277–1296 (2010).
76. Sabathe, F. & Soucaille, P. Characterization of the CipA Scaffolding Protein and In Vivo Production of a Minicellulosome in Clostridium acetobutylicum. *J Bacteriol* **185**, 1092–1096 (2003).
77. Lilly, M., Fierobe, H.-P., van Zyl, W. H. & Volschenk, H. Heterologous expression of a Clostridium minicellulosome in Saccharomyces cerevisiae. *FEMS Yeast Res* **9**, 1236–1249 (2009).
78. Anderson, T. D. *et al.* Assembly of Minicellulosomes on the Surface of Bacillus subtilis. *Appl Environ Microbiol* **77**, 4849–4858 (2011).
79. Morais, S. *et al.* Contribution of a Xylan-Binding Module to the Degradation of a Complex Cellulosic Substrate by Designer Cellulosomes. *Appl Environ Microbiol* **76**, 3787–3796 (2010).
80. Moraïs, S. *et al.* Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. *MBio* **1**, (2010).
81. Smith, S. & Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041 (2007).
82. Weissman, K. J. & Müller, R. Protein–Protein Interactions in Multienzyme Megasynthetases. *ChemBioChem* **9**, 826–848 (2008).
83. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc Natl Acad Sci USA* **96**, 1846–1851 (1999).
84. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol* **23**, 1171–1176 (2005).

85. Tang, L., Fu, H. & McDaniel, R. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chemistry & Biology* **7**, 77–84 (2000).
86. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
87. Skerker, J. *et al.* Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell* **133**, 1043–1054 (2008).
88. Park, S.-H., Zarrinpar, A. & Lim, W. A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061–1064 (2003).
89. Bashor, C. J., Helman, N. C., Yan, S. & Lim, W. A. Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics. *Science* **319**, 1539–1543 (2008).
90. Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. A. Domains, Motifs, and Scaffolds: The Role of Modular Interactions in the Evolution and Wiring of Cell Signaling Circuits. *Annu Rev Biochem* **75**, 655–680 (2006).
91. Page, C. C., Moser, C. C., Chen, X. & Dutton, P. L. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Nature* **402**, 47–52 (1999).
92. Bretschger, O. *et al.* Current Production and Metal Oxide Reduction by Shewanella oneidensis MR-1 Wild Type and Mutants. *Appl Environ Microbiol* **73**, 7003–7012 (2007).
93. Urlacher, V. B. & Eiben, S. Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol* **24**, 324–330 (2006).
94. Shiota, N., Kodama, S., Inui, H. & Ohkawa, H. Expression of human cytochromes P450 1A1 and P450 1A2 as fused enzymes with yeast NADPH-cytochrome P450 oxidoreductase in transgenic tobacco plants. *Biosci Biotechnol Biochem* **64**, 2025–2033 (2000).
95. Dodhia, V. R., Fantuzzi, A. & Gilardi, G. Engineering human cytochrome P450 enzymes into catalytically self-sufficient chimeras using molecular Lego. *J Biol Inorg Chem* **11**, 903–916 (2006).
96. Aliverti, A. & Zanetti, G. A three-domain iron-sulfur flavoprotein obtained through gene fusion of ferredoxin and ferredoxin-NADP+ reductase from spinach leaves. *Biochemistry* **36**, 14771–14777 (1997).
97. Agapakis, C. M. *et al.* Insulation of a synthetic hydrogen metabolism circuit in bacteria. *J Biol Eng* **4**, 3 (2010).
98. Ihara, M. *et al.* Light-driven hydrogen production by a hybrid complex of a [NiFe]-hydrogenase and the cyanobacterial photosystem I. *Photochem Photobiol* **82**, 676–682 (2006).

99. Lubner, C. E. *et al.* Solar hydrogen-producing bionanodevice outperforms natural photosynthesis. *Proceedings of the National Academy of Sciences* **108**, 20988–20991 (2011).
100. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* **333**, 470–474 (2011).
101. Conrado, R. J. *et al.* DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res* **40**, 1879–1889 (2012).

Figure 1

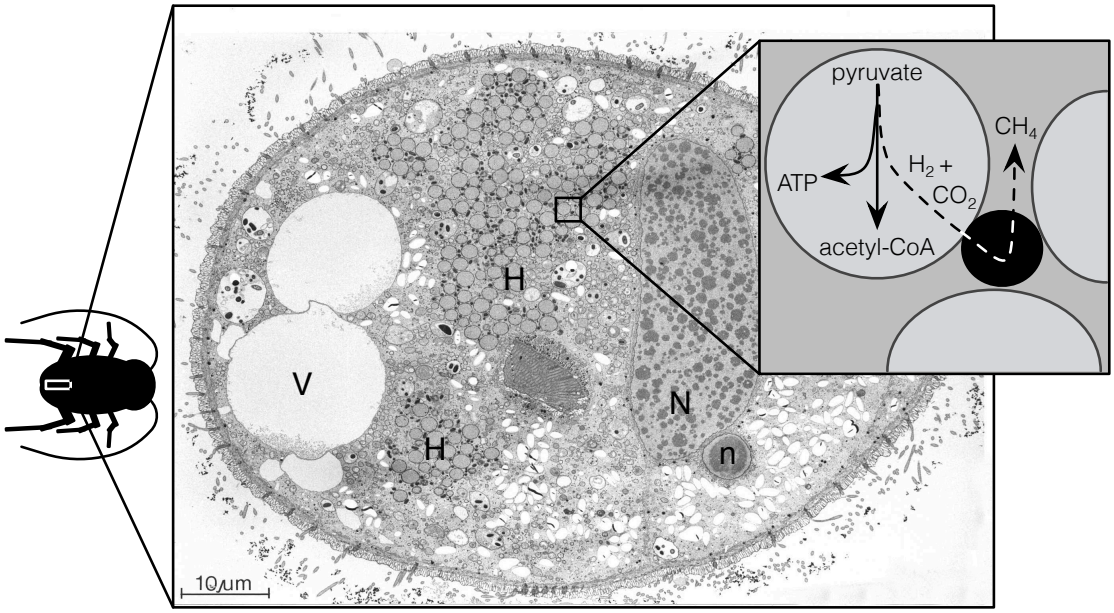


Figure 2

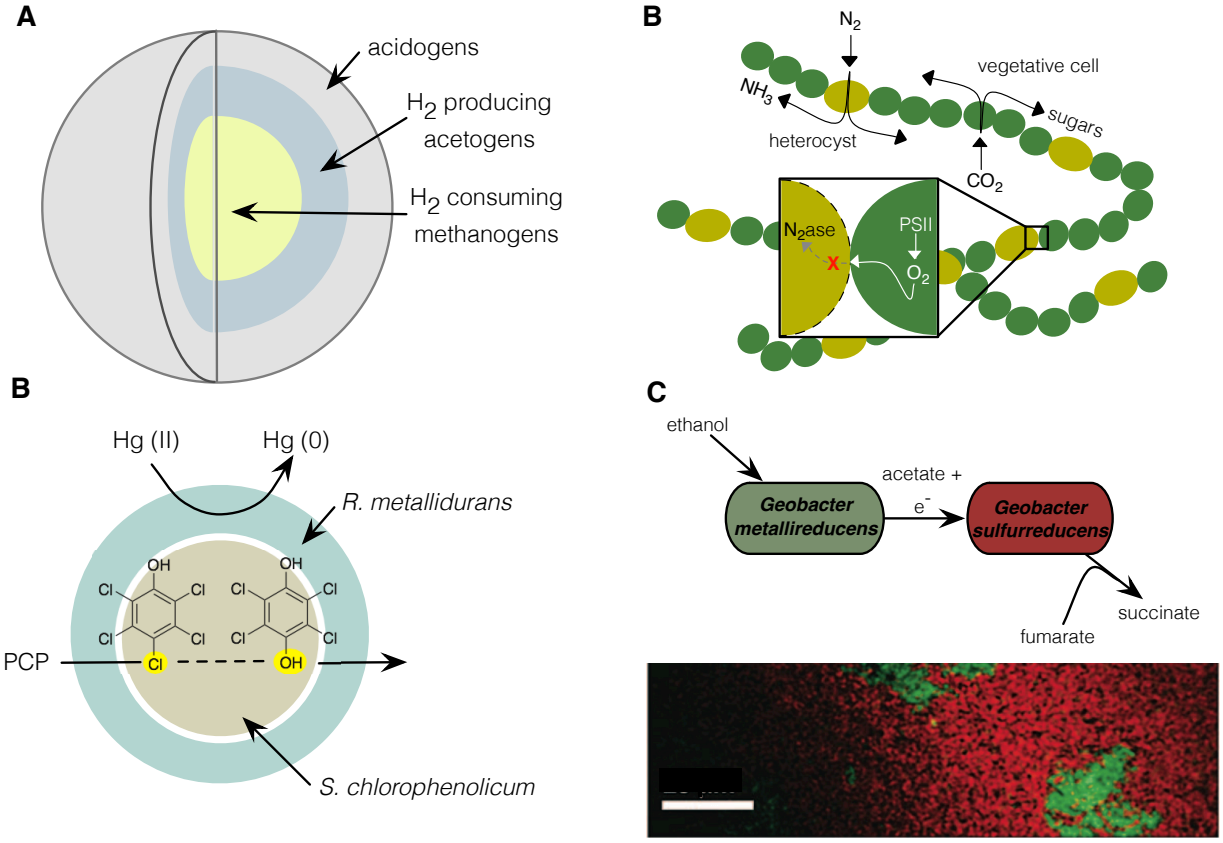


Figure 3

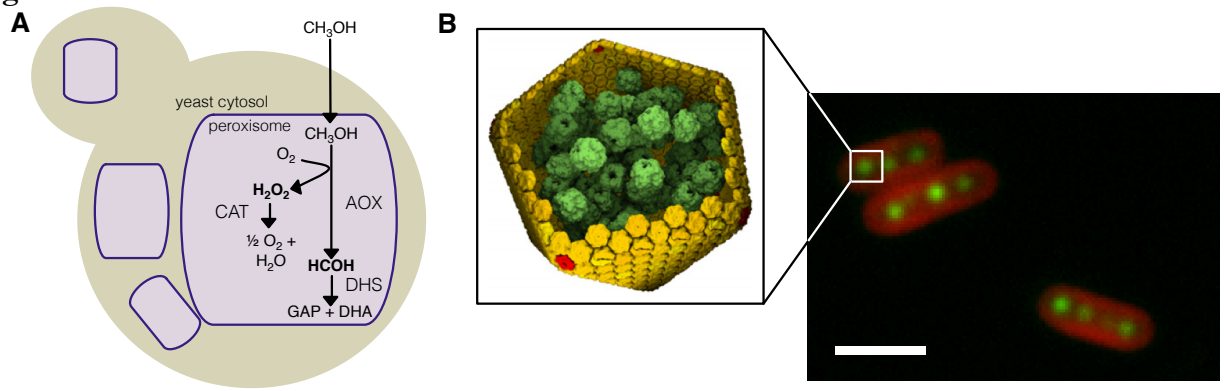


Figure 4

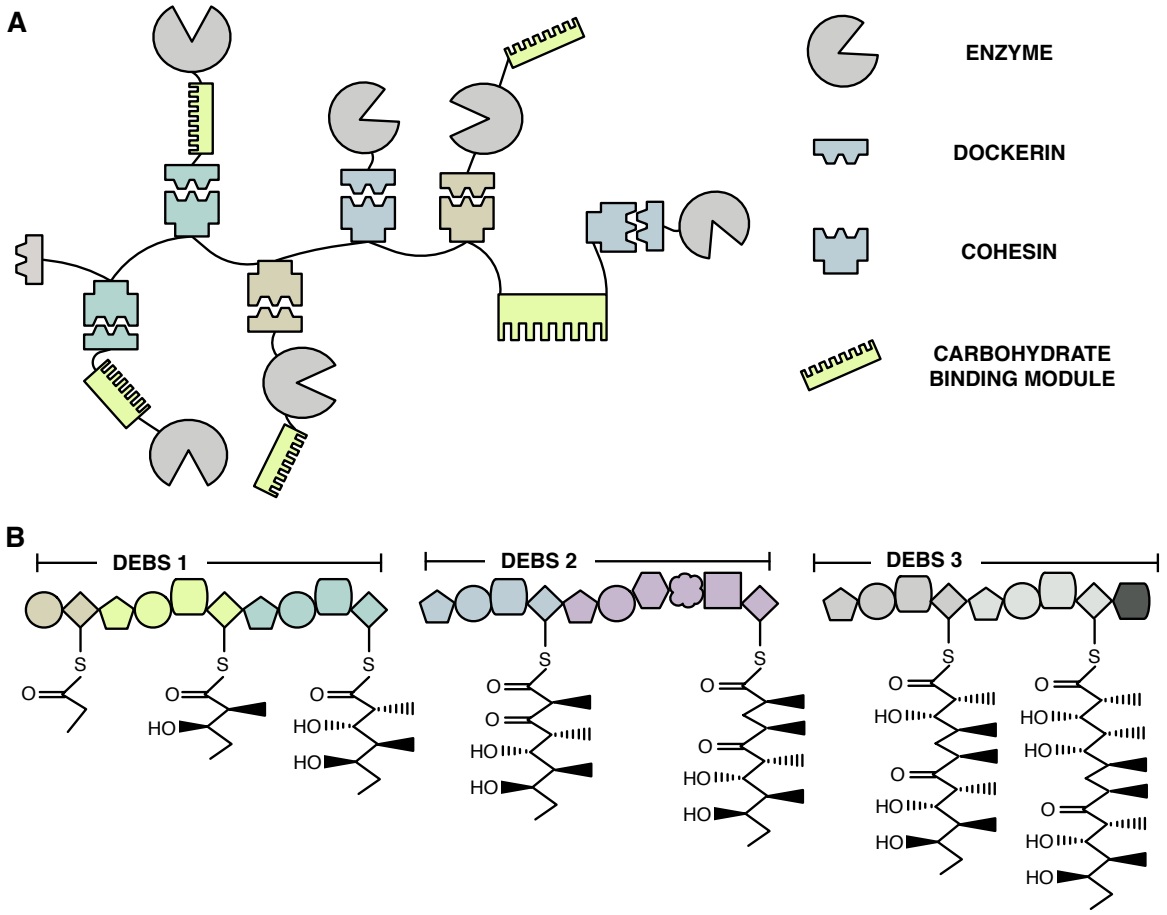
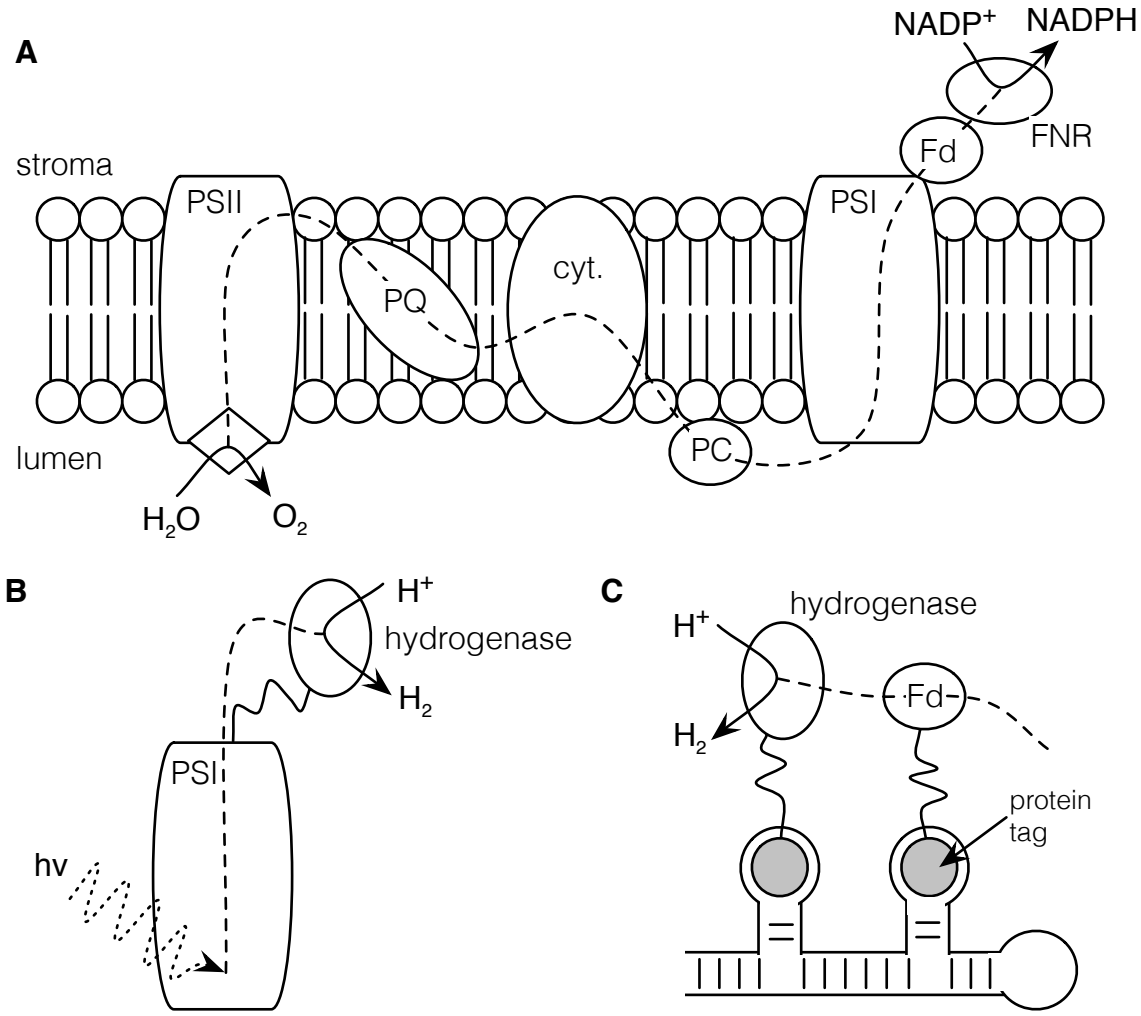


Figure 5


A
NADP+
NADPH
FNR
Fd
stroma
PSII
PSI
cyt.
PQ
PC
lumen
H2O
O2
B
H+
hydrogenase
H2
PSI
hv
C
hydrogenase
H+
H2
Fd
protein
tag